On Realistic Errors in Quantum Computers

by

Joshua Skanes-Norman

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Masters of Mathematics
in
Applied Mathematics (Quantum Information)

Waterloo, Ontario, Canada, 2021

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see the included Statement of Contributions. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

With the exception of chapter 2, this thesis has been authored by Joshua Skanes-Norman.

Chapter 2 is based on work co-authored by Joel Wallman and Joshua Skanes-Norman. This project was initiated by Joel Wallman, and Joshua Skanes-Norman later contributed new results and revised the manuscript. Specifically, the contents of sections 2.3 and 2.4 (including figure 2.1 and the code to generate it) were developed by Joshua Skanes-Norman under the supervision of Joel Wallman.

Abstract

In this thesis, we are concerned with the problem of characterizing noise associated with implementations of quantum circuits. We first explore the notion of error rates of quantum circuits and argue that the semantic distinction between process fidelity as "average error rate" and diamond distance as "worst-case error rate" is wrong. To this end, we analyze their worst-case and average analogies, and show they are proportional to their original measures. We show that the non-unital contributions to the diamond distance are negligible and the discrepancy between the process fidelity and diamond distance is primarily due to unitary errors.

We then turn to a new analysis of cycle benchmarking, a randomized benchmarking-like protocol to estimate the process fidelity of a cycle as engineered by randomized compiling. Using this approach, we prove that gate-dependent noise on the randomized gates is described by a single perturbation term that decays rapidly as long as the implementation is close to a representation. We also comment on how the protocol can be extended to qudits, what the cycle benchmarking decay actually measures, and how our analysis is amenable to the Fourier analysis of randomized benchmarking. We end with a discussion of how the gate-dependent cycle benchmarking process fidelity relates to gate-dependent randomized compiling.

Acknowledgements

I would like to thank Joseph Emerson and Joel Wallman for their supervision throughout my time at the University of Waterloo, and giving me the opportunity to explore a wide range of ideas. I would also like to thank Arnaud Carignan-Dugas for his guidance when it came to new projects, pointing me towards helpful resources, or simply giving me an impromptu presentation of some idea he was working on.

To all my friends and family near and far, thank you. I appreciate the innumerable phone calls, visits, and trips. Without those brief reprieves, I would have not been able to do this. The support throughout this time has and always will be immensely appreciated.

The work presented in this thesis was partially funded by NSERC through a CGS-M Scholarship, the province of Ontario through an Ontario Graduate Scholarship, as well the University of Waterloo.

Dedication

For those of you who have never (and also those who have).

Table of Contents

List of Figures

List of Symbols

G	A group.
$\mathbb{E}_S$	The uniform average over a set S.
ϕ	A representation of a group G over a vector space V.
ϕ_λ	The sub- or irreducible representation of ϕ over a subspace λ of V.
χ_λ	The character of a sub- or irreducible representation ϕ_λ.
$\mathsf{U}(d)$	The unitary group of dimension d.
P_q	The Pauli group on q qubits.
C_q	The Clifford group on q qubits.
$\mathcal{E}$	A quantum channel.
$\mathcal{U}$	A unitary channel.
$\epsilon_\diamond(\mathcal{E})$	The diamond distance from the identity channel to the channel $\mathcal{E}$.
$f(\mathcal{E},\psi), r(\mathcal{E},\psi)$	The ψ-(in)fidelity of a quantum channel $\mathcal{E}$.
$f_A(\mathcal{E}), r_A(\mathcal{E})$	The average (in)fidelity of a quantum channel $\mathcal{E}$.
$f_P(\mathcal{E}), r_P(\mathcal{E})$	The process (in)fidelity of a quantum channel $\mathcal{E}$.
Φ_d	A maximally entangled state on a d-dimensional system.
$f_\lambda(\mathcal{E}), r_\lambda(\mathcal{E})$	The λ-subspace (in)fidelity of a quantum channel $\mathcal{E}$.
$\tau(p,q)$	The total variation distance between probability distributions p, q.
$\mathcal{D}_\mathsf{G}$	A G-twirled quantum channel.
$\theta(G)$	A CPTP implementation of a unitary gate G.

Chapter 1

Introduction

Quantum computers and quantum information processing are poised to become efficient and effective tools for solving novel problems in disparate fields, including simulation of quantum systems [1, 2] and integer factorization [3]. Unfortunately, physical quantum systems are prone to a much wider variety of errors than their classical counterparts, and, in general, quantum computers will imprecisely implement any desired operation. For this reason it is important to understand how to extract information about errors from quantum systems and what error measures reliably predict the outcome of quantum computers. Having this knowledge will allow for the development of error mitigation schemes and quantum error correction codes that could lead to fault-tolerant quantum computing [4].

While an error or noise associated with a quantum process can be fully characterized via quantum process tomography [5] or gate set tomography [6], these protocols can prove costly as they require a number of measurements that scales exponentially with the system size — even for a small number of qubits, this is intractable. Randomized benchmarking (RB) experiments are a scalable, robust alternative to tomographic protocols that require fewer measurements to extract relevant figures of merit (e.g., the process fidelity [7] or the unitarity [8]). By cleverly sampling random quantum circuits whose components are building blocks of actual quantum computations RB allows us to infer some partial information about those components. Taking the expectation over the random circuits enacts a twirling or averaging operator, forcing the error onto particular subspaces, and so the measurement outcomes can be fit to an exponential decay when varying the length of the circuits. The decay rate will have a simple relationship to a figure of merit based on the set from which the circuits are sampled. Furthermore, RB is robust to state preparation and measurement (SPAM) errors which are contained in the constant. The building blocks

are generally individual gates, but they could also be specific combinations of gates that are frequently implemented in parallel.

The question, however, is how well the partial information of a figure of merit captures the overall behaviour of a quantum circuit. Chapter 2 explores this question. Broadly, there are two classes of measures to characterize how well a quantum circuit has been implemented on a quantum computer. The first are fidelity based measures, which consider the overlap of the action of an implemented quantum circuit on a state against that of its ideal; the process fidelity is the figure of merit that can be efficiently extracted from standard RB experiments. The second class are based on the trace distance, which has a simple relation to the difference between two probability distributions. Among its derivatives, the diamond distance is frequently used in theoretical results as it has the desirable property of being stable under tensor products, and can be directly related to the total variation distance of an ideal quantum circuit to its noisy counterpart. While it would be ideal to calculate the diamond distance directly, it is still prohibitively expensive to compute with state of the art optimization techniques.

There is an idea proliferated through the quantum information literature that the process fidelity captures the average performance of a quantum process, while the diamond distance captures the worst-case performance. Here, we consider similar but alternative fidelity-based and norm-based measures (namely, the minimum state fidelity[1] and an average version of the diamond norm) and show that, in general, they exhibit similar scalings to their "average" and "worst-case" counterparts. We then use a polar decomposition of the leading Kraus operator [9] of a quantum channel to demonstrate that the disconnect between the two measures is actually due to the square root contribution of the unitary part of a quantum channel, while there are only linear contributions from stochastic and decoherent noise. While other celebrated bounds [10] already capture this information, theorem 9 shows exactly how this factorization of an arbitrary quantum channel contributes to the bound, and is tight for both stochastic and unitary channels. We also present improved bounds on the diamond distance in terms of the process infidelity and unitarity [8], which quantifies how close a channel is to being unitary.

Chapter 3 presents a new formulation of cycle benchmarking (CB) [11] to prove its robustness to gate-dependent errors on randomizing gates. CB is an efficient RB-style protocol developed to estimate the process fidelity of cycles of gates in randomized compiling (RC) [12]. RC is a compilation technique that inserts redundant twirling gates to tailor the destructive errors associated with noisy gates into stochastic errors. Between applications of the cycle of interest, the CB algorithm inserts randomized gates in a similar

[1] Equivalently, the maximum state infidelity.

fashion to RC to replicate and amplify the errors, which can then be isolated using specific measurements. While CB enjoys many of the benefits of RB (including, but not limited to, robustness to SPAM) as well as many of its own, its somewhat original presentation made it difficult to adapt to an analysis of robustness to gate-dependence [13, 14]. Our new formulation follows an approach that directly uses results from representation theory as developed in [15], from which we prove a similar statement to [13]. We also note that this new formulation is amenable to the methods used to improve the statistical analysis and sample complexity of RB as developed in [16, 17].

In a sense, RC is the answer to the problems of unitary errors demonstrated in chapter 2 of this thesis, and as CB is used to benchmark gates in RC, it is the link between the ideas explored in this thesis. We comment on its connection to CB for gate-dependent randomizing errors in chapter 4. The rest of this chapter is dedicated to introducing concepts that will be used throughout the remainder of the thesis.

1.1 A brief introduction to quantum computing

A quantum system is described by a Hilbert space $\mathcal{H}$. For the purpose of this thesis, a quantum computer is a triple $(\Psi, \mathbb{G}, \mathbb{M})$ where Ψ is a set of states (vectors) that can be initialized on the Hilbert space $\mathcal{H}$, $\mathbb{G} \subseteq \mathsf{U}(\mathcal{H})$ is a set of unitary transformations contained within (and ideally approximating) the full unitary group, and $\mathbb{M}$ is a set of Hermitian (or self-adjoint) measurement operators. A quantum computation is the preparation of a state $\psi \in \Psi$, the application of a series of unitaries $\{U_i\}_{i=1}^n \subset \mathbb{G}$, and a measurement in the basis of an operator $M \in \mathbb{M}$, followed by a series of post-processing steps to reconstruct a quantity of interest from the measurements. We restrict to finite-dimensional Hilbert spaces ($\mathcal{H} = \mathbb{C}^d$) for which $\mathsf{U}(\mathcal{H}) = \mathsf{U}(d)$ is the group of $d \times d$ unitary matrices, and focus on qubits ($d = 2$), though some results hold for qudits as well.

A pure state ψ can be represented as a unit vector or ket

$$|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i |i\rangle \in \mathbb{C}^d \tag{1.1}$$

where $\{|i\rangle\}_{i=0}^{d-1}$ is an orthonormal basis for $\mathbb{C}^d$, dubbed the computational basis. $\alpha_i \in \mathbb{C}$ are the probability amplitudes that satisfy $\sum_{i=0}^{d-1} |\alpha_i|^2 = 1$. The dual of a ket is a bra, which is its conjugate transpose $\langle\psi| = |\psi\rangle^\dagger$.

An observable $A : \mathbb{C}^d \to \mathbb{C}^d$ is a self-adjoint operator ($A = A^\dagger$), and so has an eigenvalue decomposition

$$A = \sum_i \lambda_i \Pi_i \tag{1.2}$$

where λ_i are real numbers, Π_i is the projector onto the λ_i subspace, and the projectors satisfy the relations

$$\begin{aligned} \Pi_i \Pi_j &= \delta_{ij} \Pi_i \\ \sum_i \Pi_i &= I. \end{aligned} \tag{1.3}$$

The set $\{\Pi_i\}$ is called a projection-valued measure (PVM). We could equally insist on having a positive operator-valued measure (POVM) $\{F_i\}$ with each F_i positive semi-definite and $\sum F_i = I$, and by Naimark's dilation theorem this is equivalent to a PVM corresponding to an observable A on a larger Hilbert space.

By the Born rule, measuring the observable A when the quantum system is in the state $|\psi\rangle$ returns an eigenvalue λ_i of A with probability

$$p(i) = \langle\psi| \Pi_i |\psi\rangle = \mathrm{Tr}(|\psi\rangle\langle\psi| \Pi_i) \tag{1.4}$$

and the linearity of the trace implies $\sum_i p(i) = 1$. From 1.4, we see that the probability of measuring λ_i does not change under a global phase which is a transformation of the form $|\psi\rangle \to e^{i\theta} |\psi\rangle$. As the global phase of a state does not change measurement outcomes, we can restrict to rank-1 projectors $\psi = |\psi\rangle\langle\psi| \in \mathbb{C}^{d\times d}$ with $\mathrm{Tr}(\psi) = 1$ without loss of generality. Considering states as operators also allows us to define mixed states

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{1.5}$$

where $\sum_i p_i = 1$ and the $|\psi_i\rangle\langle\psi_i|$ are rank-1 projectors. The set of all states $\mathbb{D}_d$ is called the density operators — $\rho \in \mathbb{D}_d$ is a Hermitian, positive semi-definite operator with $\mathrm{Tr}(\rho) = 1$, and we call ρ pure if and only if $\rho = \rho^2$.

A gate set $\mathbb{G}$ is the set of unitary operations that can be performed on a quantum computer. In general, we want $\mathbb{G}$ to be a finite dense subset of the unitary group $\mathsf{U}(d)$ so we can approximate any unitary to arbitrary precision — we call such a gate set universal. This choice allows for the selection of a small number of high quality gates that can generate arbitrary elements of $\mathsf{U}(d)$. While there are many choices for $\mathbb{G}$, for q qudits of

prime dimension, a common approach is to use the generalized Clifford group $\mathsf{C}(d^q)$ and a small number of non-Clifford gates [18]. The Clifford group is the normalizer of the Weyl-Heisenberg group $\mathsf{WH}(d^q)$, that is

$$\mathsf{C}(d^q) = \{U \in \mathsf{U}(d) : U\mathsf{WH}(d^q)U^\dagger = \mathsf{WH}(d^q)\}, \tag{1.6}$$

and specifically for $d = 2$, $q = 1$, $\mathsf{WH}(2) = \mathsf{P}$, the well known Pauli group. While we give a more thorough description in section 3.1.1, the Pauli group P on one qubit is the identity matrix, the Pauli matrices

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{1.7}$$

along with a phase in $\{\pm 1, \pm i\}$. The Pauli group on q qubits P_q is the tensor product of q copies of P multiplied by a phase. We also consider the projective Pauli group is the Pauli group (the group modulo the phase) and we say an element of the Pauli group is normalized in some matrix representation if its Frobenius norm is 1. The Clifford group on q qubits C_q is generated by local Hadamard H and phase R gates on each qubit, and controlled-Z CZ (or any other controlled) gates on pairs of qubits[2]

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad R = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \quad CZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{1.8}$$

A common choice to make this set universal is $T = \sqrt{R}$.

None of these gate operations will be perfect in practice, so instead of considering unitary evolution of a quantum state we instead describe their evolution using quantum channels. A quantum channel

$$\mathcal{E} : \mathbb{D}_d \to \mathbb{D}_d \tag{1.9}$$

is a linear map that is completely positive ($\mathcal{E} \otimes \mathcal{I}_k$ is positive for all k, where $\mathcal{I}_k$ is the identity channel) and trace-preserving ($\mathrm{Tr}(\mathcal{E}(\rho)) = 1$), or CPTP. The CP constraint ensures that the the state is mapped to a positive operator under the action of $\mathcal{E}$ whether coupled to another quantum system or not. The trace-preserving constraint is used to preserve probabilities and can be relaxed to trace-non-increasing ($\mathrm{Tr}(\mathcal{E}(\rho)) \leq 1$), which can also

[2]Though not necessarily all pairs.

be viewed as tracing out the unitary dynamics on a larger system by the Stinespring factorization theorem.

The Choi matrix [19] of a quantum channel is defined as

$$\text{Choi}(\mathcal{E}) = \sum_{ij} E_{ij} \otimes \mathcal{E}(E_{ij}) \tag{1.10}$$

where $E_{ij} = e_i e_j^\dagger$ and $\{e_i\}$ is the canonical orthonormal basis of $\mathbb{C}^d$. Importantly, as $\sum_{ij} E_{ij} \otimes E_{ij}$ is positive and $\mathcal{E}$ is positive, $\text{Choi}(\mathcal{E})$ is positive semi-definite with a spectral decomposition

$$\text{Choi}(\mathcal{E}) = \sum_{i=1}^{d^2} \text{vec}(K_i)\text{vec}^\dagger(K_i) \tag{1.11}$$

where $\text{vec} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d^2}$ is the standard column stacking or vectorization map satisfying the identity $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$.

By considering different isomorphisms of $\mathbb{C}^{d\times d}$[3], the action of a quantum operation $\mathcal{E} : \mathbb{D}_d \to \mathbb{D}_d$ on a state $\rho \in \mathbb{D}$ can then be written in a Kraus operator decomposition as

$$\mathcal{E}(\rho) = \sum_{i=1}^{d^2} K_i \rho K_i^\dagger \tag{1.12}$$

where the Kraus operators K_i are the inverse vectorizations of those found in equation 1.11. The trace non-increasing constraint implies $\sum_i K_i^\dagger K_i \leq \mathcal{I}_d$, with equality if $\mathcal{E}$ is trace preserving. Kraus operators are not unique in general but are equivalent up to a unitary transform. We will follow [9] and refer to the above choice from the Choi decomposition as the canonical Kraus operators and order them according to their Frobenius norm as

$$\|K_1\|_2^2 \geq \ldots \geq \|K_{d^2}\|_2^2 \geq 0. \tag{1.13}$$

As the Kraus operators arise from a spectral decomposition and

$$\text{vec}(A)^\dagger \text{vec}(B) = \text{Tr}(A^\dagger B), \tag{1.14}$$

the Kraus operators as matrices are orthogonal with respect to the Hilbert-Schmidt inner product,

$$\text{vec}(K_i)^\dagger \text{vec}(K_j) = \text{Tr}(K_i^\dagger K_j) = \|K_i\|_2^2 \delta_{ij}. \tag{1.15}$$

[3]That is, $\mathbb{C}^{d\times d} \cong \mathbb{C}^{d^2} \cong \mathbb{C}^d \otimes \mathbb{C}^d$.

A similar construction is the Jamiołkowski-isomorphic state $J(\mathcal{E}) = (\mathcal{E} \otimes \mathcal{I})(\Phi)$ which also defines a one-to-one correspondence between states $J(\mathcal{E})$ and quantum channels $\mathcal{E}$, where $\Phi = \sum_{j,k \in \mathbb{Z}_d} |jj\rangle\langle kk|/d$ is a maximally entangled state of two systems.

In modelling errors associated with quantum errors we will consider unitary, stochastic, and decoherent channels. A unitary channel has a single Kraus operator U which is unitary. In general, this could be an under- or over-rotation of angle θ around a given axis associated with a Pauli P, that is

$$\begin{aligned} \mathcal{U}(\rho) &= U\rho U^\dagger \\ U &= e^{i\theta P} = \cos\theta I + i \sin\theta P. \end{aligned} \tag{1.16}$$

This can be relaxed to a mixed unitary channel with Kraus operators $\sqrt{p_i}U_i$ where $\{U_i\}$ is a set of unitaries and p a probability distribution. A channel is stochastic if it admits a Kraus operator decomposition that is trace-orthogonal with respect to the Hilbert-Schmidt inner product and contains a Kraus operator proportional to the identity — this implies that all other Kraus operators are traceless. Examples of stochastic channels include the depolarizing channel and Pauli channels.

A channel is decoherent if its largest (in Frobenius norm) canonical Kraus operators is close to a multiple of the identity and positive semi-definite[4]. An example is the qubit amplitude damping channel

$$\begin{aligned} \mathcal{E}(\rho) &= K_1 \rho K_1^\dagger + K_2 \rho K_2^\dagger \\ K_1 &= \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \quad K_2 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{1.17}$$

which describes a particle moving from an excited state to its ground state.

1.2 Distance measures

The output of a quantum computer can be thought of as a probability distribution dependent on the input state, a unitary circuit, and a measurement. When any of these components is noisy, the probability distribution will be perturbed. In this sense, given a noisy quantum circuit, the difference between the noisy probability distribution and

[4] A more specific definition can be found in [9], but this will suffice for the purpose of this thesis.

the ideal probability distribution is the meaningful metric to characterize how well the computation is performed. This is captured by the total variation distance

$$\tau(p,\tilde{p}) = \frac{1}{2}\sum_i |p(i) - \tilde{p}(i)| \tag{1.18}$$

where $p, \tilde{p}$ denote the ideal and noisy probability distributions respectively, with i associated with event λ_i. Throughout this thesis we are motivated by the question of how close some CPTP implementation $\tilde{\mathcal{U}}$ of a quantum gate or circuit is to its ideal $\mathcal{U}$. In particular, we want to characterize how close the CPTP error map $\mathcal{E} = \tilde{\mathcal{U}}\mathcal{U}^\dagger$ associated with $\tilde{\mathcal{U}}$ is to the identity channel $\mathcal{I}$. To build intuition about how to properly characterize these errors, we can expand the equation 1.18 using the Born rule to isolate the contribution of the implementation error to the total variation distance assuming an ideal PVM

$$\begin{aligned}
\tau(p,\tilde{p}) &= \frac{1}{2}\sum_i |p(i) - \tilde{p}(i)| \\
&= \frac{1}{2}\sum_i |\operatorname{Tr}(\Pi_i\psi) - \operatorname{Tr}(\Pi_i\rho_\psi)| \\
&= \frac{1}{2}\sum_i |\operatorname{Tr}(\Pi_i(\mathcal{I}-\mathcal{E})(\psi)|,
\end{aligned} \tag{1.19}$$

where $\rho_\psi = \mathcal{E}(\psi)$ and $\psi = \mathcal{U}(\phi)$ for some prepared state ϕ. If the pure state ψ is in the basis of the measurement, then there is some j such that $\operatorname{Tr}(\Pi_j\psi) = 1$ and for $i \neq j$ $\operatorname{Tr}(\Pi_i\psi) = 0$, so 1.19 reduces to

$$\begin{aligned}
\tau(p,\tilde{p}) &= \frac{1}{2}(1 - \operatorname{Tr}(\Pi_j\mathcal{E}(\psi)) + \sum_{j\neq i}\operatorname{Tr}(\Pi_i\mathcal{E}(\psi))) \\
&= 1 - \operatorname{Tr}(\psi\mathcal{E}(\psi))
\end{aligned} \tag{1.20}$$

where we have used $\sum_i \operatorname{Tr}(\Pi_i\mathcal{E}(\psi)) = 1$ for $\mathcal{E}$ CPTP. The quantity

$$f(\mathcal{E},\psi) = \operatorname{Tr}(\psi\mathcal{E}(\psi)) \tag{1.21}$$

is the ψ-state fidelity of $\mathcal{E}$. In most applications of quantum computing, the state and measurement will not be in the same basis, and we can instead apply a norm based bound to the total variation distance. Applying von Neumann's inequality and noting that the ideal measurement operators have a singular value of 1 gives

$$\tau(p,\tilde{p}) \leq \frac{1}{2}\sum_i \|(\mathcal{I}-\mathcal{E})(\psi)\|_1 = \frac{m}{2}\|(\mathcal{I}-\mathcal{E})(\psi)\|_1 \tag{1.22}$$

where m is the number of measurement outcomes. We could have considered the difference between probabilities of individual events occurring instead. Here, $\|A\|_1$ is the trace or Schatten 1-norm, which is equivalently the sum of the singular values of A or the trace of the operator $\sqrt{A^\dagger A}$. Dependent on the context of the quantum computation we will either consider a fidelity or a norm-based distance measure, and as demonstrated in chapter 2 we show that measuring in a slightly rotated basis induces errors to which the fidelity is not sensitive. These errors are computationally relevant for a variety of algorithms including Shor's algorithm [3], where the ideal state is not in the support of the correct factors. Note that these quantities can be related to the classical analogues for probability distributions, namely the Bhattacharya overlap and the Kolmogorov distance [20].

From the ψ-state fidelity we define other fidelity-based measures. In the quantum characterization and experimental literature, the average fidelity

$$f_{\mathrm{A}}(\mathcal{E}) = \int d\psi \, \mathrm{Tr}(\psi \mathcal{E}(\psi)) \tag{1.23}$$

is ubiquitous as it is efficient to compute with RB (as discussed in section 1.4). The integral is with respect to the Haar measure which is unitarily invariant — exploiting this we have

$$\begin{aligned} f_{\mathrm{A}}(\mathcal{E}) &= \int d\psi \, \mathrm{Tr}(U^\dagger \psi U \mathcal{E}(U^\dagger \psi U)) = \int d\psi \, \mathrm{Tr}(\psi U \mathcal{E}(U^\dagger \psi U) U^\dagger) \\ &= \int d\psi \, \mathrm{Tr}(\psi \mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^\dagger(\psi)) = f_{\mathrm{A}}(\mathcal{U}\mathcal{E}\mathcal{U}^\dagger) \end{aligned} \tag{1.24}$$

for any unitary channel $\mathcal{U}$. Analogous to the ψ-state fidelity, the average fidelity is the average probability of measuring a state in its expected eigenbasis after being acted on by $\mathcal{E}$.

The process fidelity of a channel is

$$f_{\mathrm{P}}(\mathcal{E}) = \mathrm{Tr}(\Phi_d(\mathcal{E} \otimes \mathcal{I}_d)(\Phi_d)). \tag{1.25}$$

This quantity is linearly related to the average fidelity

$$f_{\mathrm{P}}(\mathcal{E}) = \frac{(d+1) f_{\mathrm{A}}(\mathcal{E}) - 1}{d}, \tag{1.26}$$

and if $\mathcal{E}$ has a Kraus decomposition with Kraus operators $\{K_i\}$ then following [21]

$$\begin{aligned} f_{\mathrm{P}}(\mathcal{E}) &= \frac{1}{d^2}\sum_{ijkl} \langle j|\,\mathcal{E}(|k\rangle\langle l|)\,|i\rangle\,\langle j|k\rangle\,\langle i|l\rangle \\ &= \frac{1}{d^2}\sum_{ij} \langle j|\,\mathcal{E}(|j\rangle\langle i|)\,|i\rangle \\ &= \frac{1}{d^2}\sum_{ijh} \langle j|\,K_h\,|j\rangle\,\langle i|\,K_h^\dagger\,|i\rangle \\ &= \frac{1}{d^2}\sum_{h} \left(\sum_j \langle j|\,K_h\,|j\rangle\right)\left(\sum_i \langle i|\,K_h^\dagger\,|i\rangle\right) \\ &= \frac{1}{d^2}\sum_{h} \mathrm{Tr}(K_h)\,\mathrm{Tr}(K_h^\dagger) = \frac{1}{d^2}\sum_h |\,\mathrm{Tr}(K_h)|^2 \end{aligned} \tag{1.27}$$

from the trace being basis invariant. We can also construct the maximum and minimum channel fidelities

$$\begin{aligned} f_M(\mathcal{E}) &= \max_\psi f(\mathcal{E},\psi) \\ f_m(\mathcal{E}) &= \min_\psi f(\mathcal{E},\psi). \end{aligned} \tag{1.28}$$

The norm-based distance measures we consider are built from the Schatten p-norms

$$\|A\|_p = \mathrm{Tr}(|A|^p)^{1/p}, \tag{1.29}$$

which reduce to the trace norm, the Frobenius norm, and the operator norm for $p = 1, 2, \infty$ respectively. As we are interested in errors close to the identity, we consider the Schatten p-distances from the identity

$$D_p(\mathcal{E},\psi) = \frac{1}{2}\|(\mathcal{E}-\mathcal{I})(\psi)\|_p. \tag{1.30}$$

The 1-distance is the trace distance bounding the total variation distance in equation 1.22. We can then choose to average, maximize, or minimize over the input state to define a distance measure. The common choice is the diamond distance from the identity

$$\epsilon_\diamond(\mathcal{E}) = \frac{1}{2}\|\mathcal{E}-\mathcal{I}\|_\diamond, \tag{1.31}$$

where the diamond norm is

$$\|\mathcal{E}\|_\diamond = \max_{\rho \in \mathbb{D}_{d^2}} \|\mathcal{E} \otimes \mathcal{I}(\rho)\|_1. \tag{1.32}$$

While capturing the distance between probability distributions like the trace distance, the diamond distance is also stable under tensor products

$$\epsilon_\diamond(\mathcal{E}) = \epsilon_\diamond(\mathcal{E} \otimes \mathcal{I}_k) \tag{1.33}$$

for arbitrary k. By considering the action of the tensor product of the error channel with the identity, the diamond distance and other stabilized distance measures not only quantify how well the ideal operation is implemented but how much it is propagating errors throughout the entire system if it only acts on a subsystem. The process fidelity f_P is also stable under tensor products, and provides an upper and lower bound on the diamond distance

$$1 - f_\mathrm{P}(\mathcal{E}) \leq \epsilon_\diamond(\mathcal{E}) \leq d\sqrt{1 - f_\mathrm{P}(\mathcal{E})} \tag{1.34}$$

where the lower bound is saturated for stochastic channels [22]. For general noise processes that are close to the identity, the upper and lower bounds can differ by orders of magnitude.

1.3 Representation theory

In this section we introduce the representation theory used in chapter 3, and the vectorized representations of quantum mechanics. We follow the notation set forth in [15], which pulls heavily from [23]. A unitary representation ϕ of a group G on a vector space V is a group homomorphism from G to $\mathsf{U}(V)$

$$\begin{aligned} &\phi : \mathsf{G} \to \mathsf{U}(V) \\ &\phi(G_1 G_2) = \phi(G_1)\phi(G_2), \ \forall G_1, G_2 \in \mathsf{G}. \end{aligned}$$

A subrepresentation ϕ_W of ϕ is a representation of a group on a vector subspace W of V such that the restriction of ϕ to W matches ϕ_W. We then call a subrepresentation ϕ_W irreducible if it has no proper subrepresentation (proper meaning $T \subset W$ for non-empty T).

From the representation theory of finite groups (and by Maschke's theorem we can also extend to compact groups), every representation of a finite group G is a direct sum of irreducible representations

$$\phi = \oplus_\lambda \phi_\lambda^{\otimes m_\lambda}. \tag{1.35}$$

m_λ is the multiplicity of the representation ϕ_λ in the direct sum, and we call a representation multiplicity-free if $m_\lambda = 1$ for all i. Two representations $\phi_1 : \mathsf{G} \to \mathsf{U}(V_1),\ \phi_2 : \mathsf{G} \to \mathsf{U}(V_1)$ are equivalent if there exists an isomorphism $T : V_1 \to V_2$ such that

$$T\phi_1(G) = \phi_2(G)T,\ \forall G \in \mathsf{G}. \tag{1.36}$$

The ϕ-twirl of a linear map $A : V \to V$ is

$$\mathcal{T}_\phi(A) = \mathbb{E}_{G\in\mathsf{G}}\phi(G)A\phi(G)^\dagger, \tag{1.37}$$

which commutes with $\phi(G)\ \forall G \in \mathsf{G}$. By Schur's lemma, the twirl has a simple form for multiplicity-free groups:

Lemma 1 (Schur's lemma for multiplicty-free groups)**.** *Let G be a group with multiplcity-free representation $\phi = \oplus_\lambda \phi_\lambda$ over a vector space V. Then for all linear maps $A : V \to V$ the ϕ-twirl of A is*

$$\mathcal{T}_\phi(A) = \sum_\lambda \frac{\mathrm{Tr}\left(P_\lambda A\right)}{\mathrm{Tr}\left(P_\lambda\right)} P_\lambda \tag{1.38}$$

where P_λ is the projector onto the support of the irreducible representation ϕ_λ.

The character χ_ϕ of a representation ϕ of a finite group G over a vector space V is

$$\begin{aligned}\chi_\phi &: \mathsf{G} \to \mathbb{C}\\ \chi_\phi(G) &= \mathrm{Tr}\left(\phi(G)\right),\end{aligned}$$

where the trace is implicitly over V. Characters are class functions. A class function α on G is constant on the conjugacy classes of G

$$\begin{aligned}\alpha &: \mathsf{G} \to \mathbb{C}\\ \alpha(G) &= \alpha(HGH^\dagger),\ \forall G, H \in \mathsf{G},\end{aligned}$$

and the inner product of two class functions α, β is

$$\langle \alpha, \beta \rangle = \mathbb{E}_{G\in\mathsf{G}}\alpha(G)\overline{\beta(G)}.$$

The characters of the irreducible representations of G form an orthonormal basis for class functions. We can then make use of the character projection formula to isolate irreducible representations:

Lemma 2 (Character projection formula)**.** *Let* G *be a group and* ϕ *be a representation of* G *with irreducible representation* ϕ_λ*. Then*

$$\mathbb{E}_{G\in\mathsf{G}}\overline{\chi_\lambda}(G)\phi(G) = \frac{1}{|\phi_\lambda|}P_\lambda \tag{1.39}$$

where P_λ *is the projector onto all subrepresentations of* ϕ *equivalent to* ϕ_λ*.*

Throughout this thesis but especially in chapter 3 we will use vectorized representations of quantum mechanics. While used in defining the Choi matrix in equation 1.11, we can vectorize density matrices with the column stacking map so that conjugating by a unitary or applying a CPTP map is then matrix multiplication of a vector. For a unitary channel $\mathcal{U}$ and quantum channel $\mathcal{E}$, we have

$$\begin{aligned}\text{vec}(\mathcal{U}(\rho)) &= \text{vec}(U\rho U^\dagger) = \overline{U}\otimes U\text{vec}(\rho)\\ \text{vec}(\mathcal{E}(\rho)) &= \text{vec}(\sum_i K_i\rho K_i^\dagger) = \left(\sum_i \overline{K_i}\otimes K_i\right)\text{vec}(\rho)\end{aligned} \tag{1.40}$$

which follows from the identity $\text{vec}(ABC) = (C^T\otimes A)\text{vec}(B)$. In this representation we can identify $\mathcal{U}$ with $\overline{U}\otimes U$ and $\mathcal{E}$ with $\left(\sum_i \overline{K_i}\otimes K_i\right)$ and will refer to these objects as superoperators. The same procedure can again be applied to noise channels to consider operations on them, such as the twirl of an error via gates on either side of it.

The column stacking map is a specific choice of matrix basis for the Pauli-Liouville representation, which associates each element of a matrix basis to a basis element of $\mathbb{C}^n$. Following the prescription in [13], the general approach is to let $\{e_i\}_{i=1}^n$ be the canonical orthonormal basis of $\mathbb{C}^n$ and $\{B_i\}_{i=1}^{d^2}$ a trace-orthonormal basis of $\mathbb{C}^{d\times d}$, that is, $\text{Tr}(B_i^\dagger B_j) = \delta_{ij}$. We then define the linear map

$$|\cdot\rangle\rangle : \mathbb{C}^{d\times d}\to\mathbb{C}^{d^2} \tag{1.41}$$

$$A\to|A\rangle\rangle = \sum_j \text{Tr}(B_j^\dagger A)e_j \tag{1.42}$$

with adjoint map $\langle\langle A| = |A\rangle\rangle^\dagger$ that specifies the trace formula

$$\langle\langle A|B\rangle\rangle = \text{Tr}(A^\dagger B) \tag{1.43}$$

which is the Hilbert-Schmidt inner product. By considering the matrix representation of a channel $\mathcal{C} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$, we have that for any $A \in \mathbb{C}^{d\times d}$

$$\begin{aligned}|\mathcal{C}(A)\rangle\rangle &= \sum_j \mathrm{Tr}(B_j^\dagger \mathcal{C}(A)) e_j \\ &= \sum_{i,j} \mathrm{Tr}(B_j^\dagger \mathcal{C}(B_i)) \,\mathrm{Tr}(B_i^\dagger A) e_j \\ &= \sum_i |\mathcal{C}(B_i)\rangle\rangle\langle\langle B_i|A\rangle\rangle \\ &= \mathcal{C}|A\rangle\rangle, \end{aligned} \tag{1.44}$$

with $\mathcal{C} = \sum_i |\mathcal{C}(B_i)\rangle\rangle\langle\langle B_i|$ — this is the Liouville representation of a quantum channel.

We then have for two channels $\mathcal{C}_1, \mathcal{C}_2$ and vector $|A\rangle\rangle$, $|\mathcal{C}_2(\mathcal{C}_1(A))\rangle\rangle = \mathcal{C}_2\mathcal{C}_1|A\rangle\rangle$. In particular for any two unitary channels $\mathcal{U} = \phi(U), \mathcal{V} = \phi(V)$, we have

$$\phi(U)\phi(V)|A\rangle\rangle = |\mathcal{U}(\mathcal{V}(A))\rangle\rangle = |UVAV^\dagger U^\dagger\rangle\rangle = |WAW^\dagger\rangle\rangle = |\mathcal{W}(A)\rangle\rangle = \phi(W)|A\rangle\rangle \tag{1.45}$$

where $\mathcal{W} = \phi(W) = \phi(UV) = \phi(U)\phi(V)$ is also a unitary channel. This implies that the Liouville representation is a proper representation of unitary groups $\mathsf{G} \subseteq \mathsf{U}(d)$.

For $d = 2^q$ we will choose the basis of $\mathbb{C}^{d\times d}$ to be the set of normalized Paulis on q qubits where $B_1 = 2^{-q/2} I$. As this basis is Hermitian, the matrix representation of all Hermiticity-preserving maps will be real-valued, which includes CPTP maps. When referring to this choice of basis we will call the representation the PTM (Pauli-transfer matrix) representation.

1.4 Randomized benchmarking

Randomized benchmarking (RB) is a protocol used to efficiently estimated the process fidelity of quantum gates that is robust to SPAM errors. This is achieved by sampling and applying sequences of quantum gates from a benchmarking group G and fitting the output to a model that depends on the representation theory of G in some vectorized form of quantum mechanics. We will restrict to the PTM for simplicity. Originally, G was chosen to be a 2-design so that the process fidelity of a quantum channel could be efficiently estimated by sampling from a finite subgroup. This property has been relaxed to other compact subgroups that are not 2-designs or a weaker algebraic structure like an approximate 2-design [24] but the general procedure remains the same. It is desirable for

the representation of G to be multiplicity-free to simplify the fitting procedure, but this can also be relaxed as demonstrated in the appendix of [25].

Given a group G with multiplicty-free representation $\phi(G) = \oplus_\lambda \phi_\lambda(G)$, the associated G-depolarizing channel is

$$\mathcal{D}_{\mathsf{G}} = \mathbb{E}_{G\in\mathsf{G}} \phi(G)\mathcal{E}\phi(G)^\dagger = \sum_\lambda f_\lambda P_\lambda \tag{1.46}$$

of a linear map $\mathcal{E}$ where $f_\lambda = \mathrm{Tr}(P_\lambda \mathcal{E})/d_\lambda$ and $d_\lambda = \mathrm{Tr}(P_\lambda)$. This can be accomplished by twirling $\mathcal{E}$ via Schur's lemma (lemma 1.38). We also have that

$$f_{\mathrm{P}}(\mathcal{E}) = \frac{\mathrm{Tr}(\mathcal{E})}{d^2} \tag{1.47}$$

for the Liouville representation of $\mathcal{E}$ [26]. This follows from noticing that averaging a channel conjugated by a unitary over the unitary group results in a depolarizing channel. As the trace (and the process fidelity) is linear and invariant under transformations of the form $\mathcal{E} \to \phi(U)\mathcal{E}\phi(U)^\dagger$ for unitary channels $\phi(U)$, we have

$$\mathrm{Tr}(\mathcal{E}) = \mathrm{Tr}(\mathbb{E}_{G\in\mathsf{G}} \phi(G)\mathcal{E}\phi(G)^\dagger) = \mathrm{Tr}(\mathcal{D}_{\mathsf{G}}) \tag{1.48}$$

giving

$$f_{\mathrm{P}}(\mathcal{E}) = \frac{1}{d^2}\sum_\lambda d_\lambda f_\lambda. \tag{1.49}$$

We show how RB allows us to estimate all of the f_λ, giving an estimate for $f_{\mathrm{P}}(\mathcal{E})$.

We follow [15] to outline the fitting model of a general RB protocol when the ideal representation of the group is finite and multiplicity free (though a similar idea was explored for benchmarking with the dihedral group in [27]). Assuming we want to isolate $f_\lambda(\mathcal{E})$, we apply the following procedure:

1. Select n random gates $\{G_i\}_{i=0}^{n-1}$ from G.
2. Select G_n from (possibly a subgroup of) G.
3. Prepare ρ.
4. Apply the sequence of gates $G_0, \ldots, G_{n-1}$.
5. Apply $G_n G_0^\dagger \ldots G_{n-1}^\dagger$.

6. Measure Q and multiply by $\overline{\chi_\lambda}(G_n)$.

7. Repeat steps 1 through 6 sufficiently many times.

Given an implementation $\theta(G) = \phi(G)\mathcal{E}$ where $\phi(G) = \oplus_\lambda \phi_\lambda(G)$ is a representation of G with irreducible representations ϕ_λ, the expectation of the above protocol is

$$\begin{aligned} p(n,\lambda) &= \mathbb{E}_{G_n \in \mathsf{H} \subseteq \mathsf{G}, \vec{G} \in \mathsf{G}^n} \langle\langle Q | \overline{\chi_\lambda}(G_n)\phi(G_n G_0^\dagger \dots G_{n-1}^\dagger)\mathcal{E}\phi(G_n)\mathcal{E}\dots\phi(G_0)\mathcal{E}|\rho\rangle\rangle \\ &= \langle\langle Q | (\mathbb{E}_{G_n \in \mathsf{H} \subseteq \mathsf{G}} \overline{\chi_\lambda}(G_n)\phi(G_n))(\mathbb{E}_{G \in \mathsf{G}} \phi(G)^\dagger \mathcal{E}\phi(G))^n |\rho\rangle\rangle \qquad (1.50) \\ &= \langle\langle Q | P_\lambda \mathcal{D}_{\mathsf{G}}^n \mathcal{E} | \rho \rangle\rangle \\ &= A_\lambda f_\lambda^n \qquad (1.51) \end{aligned}$$

by Schur's lemma 1.38 and the character projection formula 1.39 where $A_\lambda = \langle\langle Q | P_\lambda \mathcal{E} | \rho \rangle\rangle$ is the SPAM coefficient. $p(n,\lambda)$ is referred to as a decay within the literature as $f_\lambda \in [0,1]$, and fitting $p(n,\lambda)$ to an exponential curve for varying n will allow us to extract f_λ. For n_1 and n_2 we have that

$$f_\lambda = \left(\frac{p(n_2,\lambda)}{p(n_1,\lambda)}\right)^{\frac{1}{n_2 - n_1}}. \qquad (1.52)$$

The addition of step 2 and the character projection formula is to remove the need for multi-exponential fitting if the group G has many representations — omitting these steps equation 1.50 becomes $p(n) = \sum_\lambda A_\lambda f_\lambda^n$ which is not ideal. Computing that character χ_λ can be costly for large groups as it is the trace of potentially large matrix, but this presents a trade off. If there are a lot of irreducible representations, each one corresponds to a small subspace and so the characters will be efficient to compute, but if there are few irreducible representations the number of distinct exponentials in the sum $p(n)$ will be small and easy to fit. This highlights the simplicity of the fitting model when using the Clifford group or a 2-design as they have two irreducible representations. One corresponds to the identity having quality parameter $f_I = 1$ for trace-preserving noise, and thus $p(n) = A_I + A_\sigma f_\sigma^n$ where σ is the irreducible representation corresponding to mixing unitaries.

This Markovian gate-independent error model is not realistic as gates from the same group may be implemented by different physical mechanisms with different pathologies. By relaxing the assumption to allow $\mathcal{E}_G$ to depend on the gate G being implemented, [13] showed that this introduces perturbation to the fitting model of the form

$$p(n) = A_I + A_\sigma f_\sigma^n + \epsilon(n) \qquad (1.53)$$

where under suitable assumptions on the noise maps $\mathcal{E}_G$ the perturbation term $\epsilon(n)$ decays more rapidly than the exponential of f_σ, justifying neglecting it when n is big enough. This analysis was extended in [15] to show that

$$p(n, \lambda) = A_\lambda f_\lambda^n + \epsilon(n, \lambda) \tag{1.54}$$

for multiplicity-free representations of finite groups. In this sense, RB is robust to gate-dependent noise.

In chapter 3 we utilize this representation theoretic approach to RB in order to analyze CB, and then show that it is also robust to gate-dependent noise on randomizing gates.

1.5 Randomized compiling

As demonstrated in equation 1.34 (and to be further explored in chapter 2), the diamond distance scales much less favourably with unitary errors as opposed to stochastic errors. Unitary errors are ubiquitous as they not only arise from outside sources affecting the system, but also because of imperfect quantum control. This results in over or under rotations when applying unitary transformations, or due to pulses addressing qubits they should not when applied (referred to as crosstalk).

Randomization schemes like PAREC [28] and randomized decoupling [29] are effective for the mitigation of coherent errors. By randomizing how a logical circuit is embedded in a physical system with redundant gates, these schemes prevent the build up of unitary errors along specific axes of rotation at the cost of additional circuit shots or samples. Similar techniques have been employed in the design of unitary synthesis [30] and Hamiltonian simulation [31].

The randomized compiling (RC) scheme introduced in [12] utilizes redundant twirling gates to tailor arbitrary noise in a given circuit into stochastic noise. This is accomplished by separating the gate set $\mathbb{G}$ into an easy group E, a twirling group $\mathsf{T} \subseteq \mathsf{E}$, and a hard set $\mathbb{H}$ with the constraint that $HTH^\dagger \in \mathsf{E}$ for $T \in \mathsf{G}, H \in \mathbb{H}$. The canonical breakdown is setting $\mathsf{T} = \mathsf{P}_q$, E to be the group generated by T and the phase gate R, and $\mathbb{H}$ is the set containing H, $\sqrt{R}$, and controlled-Z CZ gates.

The circuits

$$\theta(E_{n+1})\theta(H_n)\dots\theta(H_0)\theta(E_0) \tag{1.55}$$

and

$$\theta(D_{n+1})\theta(H_n)\dots\theta(H_0)\theta(D_0) \tag{1.56}$$

are logically equivalent with $D_i = T_i E_i T_{i-1}^c$, $T_i^c = H_i T_i^\dagger H_i^\dagger$, and $T_{n+1} = T_{-1}^c = I$. The D_i are referred to as dressed gates. If there are linear superoperators $\mathcal{L}, \mathcal{R}$ such that $\theta(E) = \mathcal{L}\phi(E)\mathcal{R}$ for $E \in \mathsf{E}$ for a representation $\phi(E)$, then around each hard gate H_i

$$\theta(D_i)\theta(H_{i-1})\theta(D_{i-1}) = \mathcal{L}\phi(D_i)\mathcal{R}\theta(H_{i-1})\mathcal{L}\phi(D_{i-1})\mathcal{R}.$$

Expanding D_i, D_{i-1} and averaging over the T_i

$$\mathcal{D}_{\mathrm{RC}} = \mathbb{E}_{T_{i-1}\in\mathsf{T}}\phi(T_{i-1}^\dagger)\phi(H_{i-1}^\dagger)\mathcal{R}\theta(H_{i-1})\mathcal{L}\phi(T_{i-1}).$$

Selecting $\mathsf{T} = \mathsf{P}_q$ the above is a Pauli channel by Schur's lemma

$$\begin{aligned}\mathcal{D}_{\mathrm{RC}} &= \sum_{\lambda\in\mathsf{P}_q} \mathrm{Tr}(\phi(H^\dagger)\mathcal{R}\theta(H)\mathcal{L}P_\lambda)P_\lambda \\ &= \sum_{\lambda\in\mathsf{P}_q} f_\lambda P_\lambda.\end{aligned}$$

The RC fidelity of $\theta(H)$ is then

$$f_{\mathrm{RC}}(\mathcal{E}) = f_{\mathrm{P}}(\mathcal{D}_{\mathrm{RC}}) = \frac{1}{4^n}\sum_\lambda f_\lambda \tag{1.57}$$

where the sum is over the subspaces corresponding to each normalized Pauli matrix. f_{RC} can be efficiently estimated using CB as discussed in chapter 3. As this channel has been made stochastic by RC, it saturates the lower bound of equation 1.34 so that the process fidelity exactly measures the implementation error.

Chapter 2

On error rates in quantum circuits

Currently, great experimental effort is being exerted to control quantum systems precisely enough to allow for scalable quantum error correction and to demonstrate quantum supremacy, that is, perform some task on an experimental quantum computer that is not viable on a conventional computer. For both of these efforts, it is inadequate to completely characterize every circuit component in place, as this would be more difficult then simply simulating a quantum computer directly, thus removing any possible computational advantage. Consequently, one of the goals of quantum characterization is to provide figures of merit and efficient methods of estimating those figures such that a sufficiently small figure of merit guarantees that fault-tolerant quantum computation is possible and/or that the total error in a circuit is sufficiently small.

There are several different metrics for quantifying the error in quantum gates that are relevant to different tasks [20]. Proofs in quantum information theory and fault-tolerant quantum computation typically appeal to the diamond distance from the identity [32, 33, 34]. There is no known method for efficiently estimating the diamond distance.

A more common experimental figure of merit is the process (equivalently, average) infidelity, because it can be efficiently and robustly measured by randomized benchmarking [7, 35, 36, 37, 22]. Upper and lower bounds on the diamond distance can be obtained from the fidelity [38]. For small error rates, the upper and lower bounds differ by orders of magnitude and cannot be substantially improved [39]. There is a belief that the infidelity captures the average "computationally relevant" error [36] and that the loose relation between the fidelity and the diamond norm is simply a difference between "average" and "worst-case" performance—this belief is still prevalent in the literature, e.g. as seen in a recent preprint [40].

In this chapter, we obtain upper- and lower-bounds on the diamond distance for general noise in terms of functions that can be efficiently estimated, namely, the infidelity and the unitarity [8], that differ by essentially a factor of $d^{3/2}$. The bounds here improve upon similar bounds obtained for unital noise in Ref. [41] by a dimensional factor and hold for general noise. To obtain these bounds, we also prove that a "worst-case" or maximum version of the infidelity is proportional to the process infidelity and that the contribution of the non-unital portion to the diamond distance due to generalized amplitude processes is at most linear in the infidelity, and so does not introduce any significant difference between the infidelity and the diamond distance. We then obtain upper- and lower-bounds on the diamond distance that scale as $\sqrt{r_{\mathrm{P}}}$ instead of r_{P} if and only if there are significant coherent contributions to the noise as quantified by the unitarity [8] (theorem 6). This is further justified by using a leading Kraus approximation [9] to bound the diamond distance in terms of the maximum infidelity, where the dominant source of error comes from unitary contributions.

Non-stochastic noise will generically introduce coherence between the expected state of the system and its orthogonal subspace, which results in some level of indeterminacy in the state after applying a quantum process. This situation results in a basis mismatch during measurement, and so it is not immediately clear whether or not infidelity is the correct choice to quantify the error. Furthermore, in general applications of quantum information, such as quantum computing, the measurement outcome is often expected to be probabilistic even in the absence of noise, that is, measurements are expected to be in a basis that does not contain the ideal state of the system[1]. Consequently, the coherent contributions to the diamond and induced trace distances will be relevant.

We provide numerical evidence to substantiate this claim. We show that when measuring and preparing states in the same basis with some small basis mismatch the total error in a quantum circuit does not generically scale with the average infidelity, but instead scales with either the infidelity or the diamond distance dependent on the magnitude of these quantities and the severity of the basis mismatch. Furthermore, if the basis mismatch is small the diamond distance can still be the dominating quantity, and so the more accurate measure.

[1]For example, this arises in algorithms based on quantum phase estimation [42] when the phase to be estimated multiplied by a power of 2 is not an integer.

2.1 Preliminaries

We will make use of the Pauli-Liouville representation of a quantum channel

$$\mathcal{E} = \begin{bmatrix} S(\mathcal{E}) & \mathcal{E}_{sdl} \\ \mathcal{E}_n & \mathcal{E}_u \end{bmatrix}, \tag{2.1}$$

which follows from equation 1.44, and we can characterize $\mathcal{E}$ based on how it differs from $\mathcal{I}$ on each of these blocks. The survival rate $S(\mathcal{E}) \in [0,1]$ accounts for the loss of trace induced by the channel, the nonunital vector $\mathcal{E}_n \in \mathbb{C}^{d^2-1}$ describes how much the channel distorts the identity, the unital block $\mathcal{E}_u \in \mathbb{C}^{d^2-1 \times d^2-1}$ represents the unitary action of the channel, and finally $\mathcal{E}_{sdl} \in \mathbb{C}^{1 \times d^2-1}$ is the state-dependent leakage. If $\mathcal{E}$ is CPTP then $S(\mathcal{E}) = 1$, which we will assume throughout this chapter, and are only concerned with channels with $\mathcal{E}_{sdl} = 0$.

The average gate infidelity is the error rate reported in randomized benchmarking experiments [37]. However, it is linearly related to the process infidelity by $r_{\mathrm{P}}(\mathcal{E}) = \frac{d+1}{d} r_{\mathrm{A}}(\mathcal{E})$ and $\mathrm{Tr}(\mathcal{E})$ by equation 1.47, where the dimensional factors are substantially more convenient for the process infidelity. Consequently, we will state all results in terms of the process infidelity. We also quantify the coherent part of a noise process using the unitarity [8]

$$\begin{aligned} u(\mathcal{E}) &= \frac{d}{d-1} \int d\psi \, \mathrm{Tr}(\mathcal{E}(\psi)^\dagger \mathcal{E}(\psi)) \\ &= \frac{1}{d^2-1} \mathrm{Tr}(\mathcal{E}_u^\dagger \mathcal{E}_u). \end{aligned} \tag{2.2}$$

Importantly, $u(\mathcal{E}) < 1$ unless $\mathcal{E}$ is unitary. The non-unital or decoherent part of a channel $\mathcal{E}$ can be quantified by the non-unital vector $\mathcal{E}_n$ and the non-unitality

$$\begin{aligned} \alpha(\mathcal{E}) &= \frac{1}{d} \|\mathcal{E}(I_d) - I_d\|_2 \\ &= \frac{1}{d} \|\mathcal{E}_n\|. \end{aligned} \tag{2.3}$$

2.2 Analyzing the discrepancy between process infidelity and diamond norm

We are primarily interested in explaining the difference between the process infidelity and the diamond distance, and determining which is more relevant dependent on context. A

common explanation appeals to the different dependence on the input state. The process infidelity is equivalent to the state infidelity averaged over input states acting on the system, whereas the diamond distance is maximized over input states acting on a system and an ancilla. As we now show, this difference can explain the dimensional factor, but not the square root. For any $d \in \mathbb{N}$, orthonormal basis $\{|j\rangle : j \in \mathbb{Z}_d\}$ of $\mathbb{C}^d$, and CPTP map $\mathcal{E}$, the process infidelity satisfies [43]

$$\frac{1}{d}\sum_{j\in\mathbb{Z}_d} r(\mathcal{E}, |j\rangle\langle j|) \leq r_{\mathrm{P}}(\mathcal{E}). \tag{2.4}$$

Consequently, the worst-case infidelity can exhibit the same dimensional scaling as the diamond norm but is always proportional to the process infidelity.

Proposition 3. *For any CPTP map $\mathcal{E}$, the worst-case infidelity satisfies*

$$\frac{d}{d+1} r_{\mathrm{P}}(\mathcal{E}) \leq r_{\mathrm{M}}(\mathcal{E}) \leq d r_{\mathrm{P}}(\mathcal{E}). \tag{2.5}$$

Furthermore, there exist CPTP maps such that

$$r_{\mathrm{M}}(\mathcal{E}) \geq \frac{d^2 r_{\mathrm{P}}(\mathcal{E})}{4(d-1)}. \tag{2.6}$$

Proof. The lower bound follows from the relation between the process infidelity and the non-negativity of the ψ-infidelity. The upper bound follows from eq. (2.4) with the non-negativity of the ψ-infidelity. Equation (2.6) can be obtained by setting $\mathcal{E} = \mathcal{U}_\phi$ where $U_\phi = I_d + (e^{i\phi} - 1)\,|0\rangle\langle 0|$, which has process infidelity

$$r_{\mathrm{P}}(\mathcal{E}) = 1 - \frac{|\operatorname{Tr} U_\phi|^2}{d^2} = \frac{2(d-1)(1-\cos(\phi))}{d^2}. \tag{2.7}$$

Evaluating the ψ-infidelity for $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ gives

$$\begin{aligned} r_{\mathrm{M}}(\mathcal{U}_\phi) \geq r(\mathcal{U}_\phi, |+\rangle\langle +|) &= \frac{1-\cos(\phi)}{2} \\ &= \frac{d^2 r_{\mathrm{P}}(\mathcal{U}_\phi)}{4(d-1)}. \end{aligned} \tag{2.8}$$

□

Proposition 3 demonstrates that the potential discrepancy between the $O(\sqrt{r_{\mathrm{P}}})$ scaling of the diamond distance and the process infidelity is not simply the difference between "average" and "worst-case". The discrepancy only arises for channels that are not stochastic. The primary examples of non-stochastic processes are relaxation and coherent processes, whose contributions account for the difference between the diamond norm and the infidelity.

Relaxation to a ground state is a common physical process that cannot be described as a stochastic channel. The canonical example is the single qubit amplitude damping channel as in equation 1.17. For multiple qubits, the energy eigenbasis (including the ground state) will generically be entangled relative to the computational basis after such a process, and so will not be aligned with the measurement basis. To quantify this non-unital behavior of a noise process, we consider the purity of the maximally mixed state after the noise process is applied, with the known identity component subtracted off. While the following proposition is stated in terms of the Schatten 2-norm, using the 1-norm instead only introduces a dimensional factor as we are dealing with a finite-dimensional space. Note that the following bound coincides with [38, Prop. 12] for $d = 2$ without a restriction on $r_{\mathrm{P}}(\mathcal{E})$.

Theorem 4. *For any CPTP map $\mathcal{E}$ we have that*

$$\alpha(\mathcal{E}) = \|\mathcal{E}(\tfrac{1}{d}I_d) - \tfrac{1}{d}I_d\|_2 \leq \sqrt{2}r_{\mathrm{P}}(\mathcal{E}). \tag{2.9}$$

Proof. For any trace-preserving map $\mathcal{E}$, we have

$$\alpha(\mathcal{E})^2 = \operatorname{Tr}\mathcal{E}(\tfrac{1}{d}I_d)^2 - \frac{2}{d}\operatorname{Tr}\mathcal{E}(\tfrac{1}{d}I_d) + \frac{1}{d} = \frac{1}{d^2}\operatorname{Tr}\mathcal{E}(I_d)^2 - \frac{1}{d}. \tag{2.10}$$

We want to relate the purity to the state fidelities averaged over a basis. Let $\{|j\rangle : j \in \mathbb{Z}_d\}$ be an eigenbasis of $\mathcal{E}(I_d)$ with corresponding eigenvalues $\lambda_j = 1 - \delta_j$. We then have

$$\begin{aligned}
\alpha(\mathcal{E})^2 &= \frac{1}{d^2}\sum_j \lambda_j^2 - \frac{1}{d} \\
&= \frac{1}{d^2}\sum_j \delta_j^2 - 2\delta_j \\
&= \frac{1}{d^2}\sum_j \delta_j^2
\end{aligned} \tag{2.11}$$

where we use

$$\sum_j \delta_j = d - \operatorname{Tr} \mathcal{E}(I_d) = 0 \tag{2.12}$$

for trace-preserving maps. We can relate the δ_j to the state fidelities by noting that

$$\begin{aligned}
\delta_j &= 1 - \langle j| \mathcal{E}(I_d) |j\rangle \\
&= r(\mathcal{E}, |j\rangle\langle j|) - \langle j| \mathcal{E}(I_d - |j\rangle\langle j|) |j\rangle \\
&= r(\mathcal{E}, |j\rangle\langle j|) - v_j,
\end{aligned} \tag{2.13}$$

using eq. (1.21), where we implicitly define v_j. The δ_j terms represent the net flow out of the state $|j\rangle\langle j|$ and satisfy

$$\begin{aligned}
\sum_j v_j &= \sum_j \sum_{k \neq j} \langle j| \mathcal{E}(|k\rangle\langle k|) |j\rangle \\
&= \sum_k \sum_{j \neq k} \langle j| \mathcal{E}(|k\rangle\langle k|) |j\rangle \\
&= \sum_k 1 - \langle k| \mathcal{E}(|k\rangle\langle k|) |k\rangle \\
&= \sum_j r(\mathcal{E}, |j\rangle\langle j|).
\end{aligned} \tag{2.14}$$

Therefore, noting that $v_j, r(\mathcal{E}, |j\rangle\langle j|) \geq 0$, we have

$$\begin{aligned}
\alpha(\mathcal{E})^2 &\leq \frac{1}{d^2} \sum_j [r(\mathcal{E}, |j\rangle\langle j|)^2 + v_j^2] \\
&\leq \frac{1}{d^2} \Big(\sum_j r(\mathcal{E}, |j\rangle\langle j| \Big)^2 + \frac{1}{d^2} \Big(\sum_j v_j \Big)^2 \\
&= \frac{2}{d^2} \Big(\sum_j r(\mathcal{E}, |j\rangle\langle j|) \Big)^2 \\
&\leq 2 r_{\mathrm{P}}(\mathcal{E})^2,
\end{aligned} \tag{2.15}$$

where the final inequality follows from eq. (2.4). □

We now obtain lower- and upper-bounds on the diamond norm of a general linear map $\mathcal{T}$ in terms of the purity of the Jamiołkowski-isomorphic state that differ by a factor of

$d^{3/2}$. We will then state upper and lower bounds on the diamond distance in terms of the infidelity and the unitarity. In particular, $u(\mathcal{E}) \in [p(\mathcal{E})^2, 1]$ with $u - p(\mathcal{E})^2 \in O(r_{\mathrm{P}}^2(\mathcal{E}))$ giving a diamond distance that scales linearly with $r_{\mathrm{P}}(\mathcal{E})$ if $\mathcal{E}$ is stochastic where $p(\mathcal{E}) = 1 - d^2 r_{\mathrm{P}}(\mathcal{E})/(d^2-1)$ is the randomized benchmarking decay constant.

Theorem 5. *For any linear map $\mathcal{T} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$,*

$$\|J(\mathcal{T})\|_2 \leq |||\mathcal{T}|||_1 \leq d^{3/2}\|J(\mathcal{T})\|_2 \tag{2.16}$$

where

$$|||\mathcal{T}|||_1 = \sup_{A\in\mathbb{C}^{d^2\times d^2}:\|A\|_1=1} \|\mathcal{T}\otimes\mathcal{I}_d(A)\|_1. \tag{2.17}$$

Proof. By [44, Thm. 6],

$$|||\mathcal{T}|||_1 = \sup_{\rho,\sigma\in\in\mathbb{D}_d} d\|(I_d\otimes\sqrt{\rho})J(\mathcal{T})(I_d\otimes\sqrt{\sigma})\|_1 \tag{2.18}$$

noting that our convention of $J(\mathcal{T})$ differs from Ref. [44] by a factor of d. The lower bound can be obtained by setting $\rho = \sigma = I_d/d$ and applying eq. (2.25). By Hölder's inequality,

$$\|ABC\|_1 \leq \|A\|_\infty\|BC\|_1 \leq \|A\|_\infty\|B\|_2\|C\|_2 \tag{2.19}$$

for arbitrary $A, B, C \in \mathbb{C}^{m\times m}$. As $\rho, \sigma \in \mathbb{D}_d$, they are positive semi-definite and have unit trace and so $\|I_d\otimes\sqrt{\rho}\|_\infty \leq 1$ and $\|I_d\otimes\sqrt{\sigma}\|_2 \leq \sqrt{d}$ for all $\rho, \sigma \in \mathbb{D}_d$. (Note that the bound in Ref. [38] uses $\|BC\|_1 \leq \|B\|_1\|C\|_\infty$ instead, which, together with the Fuchs-van de Graaf inequality [10], gives the bound in eq. (1.34).) □

Corollary 6. *For any quantum channel $\mathcal{E}$,*

$$\frac{1}{\sqrt{2}}\|J(\Delta)\|_2 \leq \epsilon_\diamond(\mathcal{E}) \leq \frac{d^{3/2}}{2}\|J(\Delta)\|_2 \tag{2.20}$$

where $\Delta = \mathcal{E} - \mathcal{I}$. In terms of the infidelity and unitarity,

$$\frac{C}{\sqrt{2}} \leq \epsilon_\diamond(\mathcal{E}) \leq d\sqrt{\frac{dC^2}{4} + \frac{r_{\mathrm{P}}(\mathcal{E})^2}{2}}, \tag{2.21}$$

where

$$C^2 = \frac{d^2-1}{d^2}(u(\mathcal{E}) - 2p(\mathcal{E}) + 1). \tag{2.22}$$

Proof. First note the factor of $1/2$ from the definition of $\epsilon_\diamond(\mathcal{E})$ in eq. (1.34) and we obtain the $\sqrt{2}$ improvement from theorem 7 for trace-preserving maps $\mathcal{E}$, so that $\operatorname{Tr}\Delta(A) = 0$ for all A. By [8, Prop. 9],

$$\begin{aligned}\|J(\Delta)\|_2^2 &= \frac{d^2-1}{d^2}u(\Delta) + \frac{\alpha(\mathcal{E})^2}{d} \\ &= \frac{d^2-1}{d^2}(u(\mathcal{E}) - 2p(\mathcal{E}) + 1) + \frac{\alpha(\mathcal{E})^2}{d}\end{aligned} \tag{2.23}$$

where we have used $\operatorname{Tr}\Delta(A) = 0$ for all A as $\mathcal{E}$ is a trace-preserving map, and $\alpha(\mathcal{E})$ is defined as in equation 2.3. The lower and upper bounds follow from the non-negativity of norms and theorem 4 respectively. □

Lemma 7. *For any traceless Hermitian matrix $M \in \mathbb{C}^{d\times d}$,*

$$\sqrt{2}\|M\|_2 \leq \|M\|_1 \leq \sqrt{d}\|M\|_2. \tag{2.24}$$

Moreover, both these bounds are saturated.

Proof. For any Hermitian matrix M, writing $M = U\boldsymbol{\eta}U^\dagger$ where $\boldsymbol{\eta}$ is a diagonal matrix whose entries are the eigenvalues $\{\eta_j\}$ of M and using the unitary invariance of the trace and Frobenius norms gives the standard bounds

$$\|M\|_2 = \sqrt{\sum_{j=1}^d \eta_j^2} \leq \|M\|_1 = \sum_{j=1}^d |\eta_j| \leq \sqrt{d\sum_{j=1}^d \eta_j^2} = \sqrt{d}\|M\|_2 \tag{2.25}$$

by the Cauchy-Schwarz inequality.

To obtain a sharper lower bound for traceless Hermitian matrices, let

$$\boldsymbol{\eta} = \boldsymbol{\eta}^+ \oplus -\boldsymbol{\eta}^- = \begin{pmatrix} \boldsymbol{\eta}^+ & 0 \\ 0 & -\boldsymbol{\eta}^- \end{pmatrix} \tag{2.26}$$

where $\oplus$ denotes the matrix direct sum and $\eta^\pm$ are both positive semi-definite. Then

$$\|M\|_2 = \sqrt{\sum_{j=1}^d \eta_j^2} = \sqrt{\|\boldsymbol{\eta}^+\|_2^2 + \|\boldsymbol{\eta}^-\|_2^2} \leq \sqrt{\|\boldsymbol{\eta}^+\|_1^2 + \|\boldsymbol{\eta}^-\|_1^2} = \sqrt{\frac{\|\boldsymbol{\eta}\|_1^2}{2}} = \frac{1}{\sqrt{2}}\|M\|_1, \tag{2.27}$$

where we have used $\|\boldsymbol{\eta}^\pm\|_1 = \frac{1}{2}\|\boldsymbol{\eta}\|_1$ which holds for traceless matrices. The above lower bound is saturated when $\eta_{11} = 1$, $\eta_{22} = -1$ and all other eigenvalues are zero. □

While the diamond distance can be interpreted as a bound on the maximum distance between the ideal and actual distributions of the output of a component of a quantum circuit [20], the performance of an individual gate within the circuit will generically be better. To quantify the typical behavior, we define an average analogy to the diamond distance.

For a non-trivial circuit, that is, $G_{k:1} \neq I_d$, M_k ($\approx \tilde{M}_k$) and ψ_k will be in different bases that are fixed relative to each other and so contribution to the error from each individual gate will typically be close to

$$t_{\text{avg},m}(\mathcal{E}) = \mathbb{E}_{\psi,\mathbf{M}|\operatorname{Tr} M=m,M^2=M} |\langle M, \mathcal{E}(\psi) - \psi\rangle|, \tag{2.28}$$

where $m = \operatorname{Tr} M_0$. We now prove that this quantity is proportional to the diamond distance, up to a dimensional factor and a factor that depends on the rank of M_0. For many problems, such as Shor's factoring algorithm, and for the measurements between rounds of error correction, $m \in O(d^\alpha)$ and so the dimensional scaling between the following upper and lower bounds is primarily a consequence of the corresponding dimensional factors in theorem 6.

Theorem 8. *For any quantum channel $\mathcal{E}$ and $m \in \mathbb{N}$,*

$$\frac{2m(m+1)\epsilon_\diamond(\mathcal{E})}{d^3(d+1)^2} \leq t_{\text{avg},m}(\mathcal{E}) \leq 2\epsilon_\diamond(\mathcal{E}). \tag{2.29}$$

Proof. The upper bound is trivial from the definition of $\epsilon_\diamond(\mathcal{E})$ and theorem 6 (although note the factor of 2). For the lower bound, let

$$t(M, \mathcal{E}, \psi) = |\langle M, \Delta(\psi)\rangle| \tag{2.30}$$

where $\Delta = \mathcal{E} - \mathcal{I}_d$, so that $t_{\text{avg},m}(\mathcal{E}) = \mathbb{E}_{\psi,M|\operatorname{Tr} M=m} t(M, \mathcal{E}, \psi)$. The absolute value in eq. (2.30) makes evaluating the mean difficult. To circumvent this, we use the identity $a^2 = a \otimes a$ for $a \in \mathbb{R}^d$ and the distributivity of the tensor product to obtain

$$\begin{aligned} \mathbb{V} &= \mathbb{E}_{\psi,M|\operatorname{Tr} M=m,M^2=M}[t(M, \mathcal{E}, \psi)^2] \\ &= \int_{\mathbb{S}_d} \mathrm{d}\psi \int_{\mathrm{U}(d)} \mathrm{d}U \operatorname{Tr}\left[\mathcal{U}(M^{(m)})^{\otimes 2} \Delta^{\otimes 2}(\psi^{\otimes 2})\right]. \end{aligned} \tag{2.31}$$

To evaluate the integrals, let $S = \sum_{i,j} |ij\rangle\langle ji|$ be the two-qudit swap gate and let $\pi_{a/s} = (\mathbb{I}_{d^2} \pm S_d)/2$ be the projectors onto the symmetric and antisymmetric subspaces of $\mathbb{C}^{d^2}$ respectively. For any Hermitian matrix M,

$$\int \mathrm{d}U \mathcal{U}^{\otimes 2}(M) = \frac{\operatorname{Tr} \pi_s M}{\operatorname{Tr} \pi_s} \pi_s + \frac{\operatorname{Tr} \pi_a M}{\operatorname{Tr} \pi_a} \pi_a \tag{2.32}$$

by Schur-Weyl duality and so, with $\operatorname{Tr} S(A \otimes B) = \operatorname{Tr} AB$,

$$\begin{aligned} \int \mathrm{d}\psi \, \psi^{\otimes 2} &= \int \mathrm{d}U \, \mathcal{U}(\phi^{\otimes 2}) \\ &= \frac{\mathbb{I}_{d^2} + S}{d^2 + d} \\ \int_{\mathrm{U}(d)} \mathrm{d}U \, \mathcal{U}(M^{(m)})^{\otimes 2} &= \frac{2m^2 + 2m}{d^2 + d} \pi_s + \frac{2m^2 - 2m}{d^2 - d} \pi_a \end{aligned} \tag{2.33}$$

where ϕ is an arbitrary pure state. Substituting this into eq. (2.31) gives

$$\begin{aligned} \mathbb{V} &= \frac{m^2 + m}{(d^2 + d)^2} \operatorname{Tr}\left[(\mathbb{I}_{d^2} + S)\Delta^{\otimes 2}(\mathbb{I}_{d^2} + S)\right] \\ &= \frac{m^2 + m}{(d^2 + d)^2} \operatorname{Tr}\left[S\Delta^{\otimes 2}(\mathbb{I}_{d^2} + S)\right] \\ &= \frac{m^2 + m}{(d+1)^2} (\|J(\Delta)\|_2^2 + \|\Delta(I_d)\|_2^2) \\ &\geq \frac{4m(m+1)\epsilon_\diamond(\mathcal{E})^2}{d^3(d+1)^2} \end{aligned} \tag{2.34}$$

where we have used $\pi_a \Delta^{\otimes 2}(\pi_s) = 0$, [8, Prop. 3], the non-negativity of norms, and theorem 6. Now note that

$$\max t(M, \mathcal{E}, \psi) \leq 2\epsilon_\diamond(\mathcal{E}), \tag{2.35}$$

and so, with theorem 6 and the non-negativity of norms,

$$\begin{aligned} t_{\mathrm{avg}}(\mathcal{E}) &\geq \frac{\mathbb{E}_{\psi,\mathbf{M}}[t(M, \mathcal{E}, \psi)^2]}{\max t(M, \mathcal{E}, \psi)} \\ &\geq \frac{2m(m+1)\epsilon_\diamond(\mathcal{E})}{d^3(d+1)^2}. \end{aligned} \tag{2.36}$$

□

2.3 A leading Kraus approximation of the diamond distance

To demonstrate the scaling of the diamond distance, we now derive an upper bound on the Schatten 1-norm of the difference between the action of a CPTP map $\mathcal{E}$ on a state and

the state itself. We are particularly interested in the case where $\mathcal{E}$ is close to the identity. In principle, there should then be a Kraus decomposition such that one of the operators is close to the identity and the rest are negligible — this is similar to the leading Kraus approximation [9]. As the diamond distance is the maximum sum of singular values of the difference between two operators acting on a state, we show that it can be bound by the singular values of its dominant coherent and decoherent contributions.

Theorem 9. *Let $\{K_j\}$ be a Kraus operator decomposition of a CPTP map $\mathcal{E}$, $K_1 = DU$ be a polar decomposition of one of the Kraus operators with singular values satisfying $\sigma_{\max}(K_1) \leq \min\{3\sigma_{\min}(K_1), 1\}$. Then for any positive integer k,*

$$\max_{\rho \in \mathbb{D}_{dk}} \tfrac{1}{2}\|\mathcal{E} \otimes \mathcal{I}_k(\rho) - \rho\|_1 \leq r_{\mathrm{M}}(\mathcal{D}) + \sqrt{r_{\mathrm{M}}(\mathcal{U})}. \tag{2.37}$$

Proof. Without loss of generality, we set $k = 1$, using the fact that $\{K_j \otimes I_k\}$ are Kraus operators of $\mathcal{E} \otimes \mathcal{I}_k$ for any k. Let $\rho \in \mathbb{D}_d$ be an arbitrary state and $\rho' = \mathcal{U}(\rho) = U\rho U^\dagger$. By the triangle inequality, we have

$$\begin{aligned}\|\mathcal{E}(\rho) - \rho\|_1 &\leq \|K_1\rho K_1^\dagger - \rho' + \rho' - \rho\|_1 + \sum_{j>1} \|K_j \rho K_j^\dagger\|_1 \\ &\leq \|\rho' - \rho\| + \|D\rho' D^\dagger - \rho'\|_1 + \sum_{j>1} \|K_j \rho K_j^\dagger\|_1. \end{aligned} \tag{2.38}$$

We will obtain upper bounds on the three terms on the right-hand-side of eq. (2.38) separately.

By convexity, $\|\rho' - \rho\|_1$ will be maximized when ρ is a pure state ψ. By the Fuchs-van de Graaf inequality [10], for any pure state ψ we have

$$\|U\psi U^\dagger - \psi\|_1 = 2\sqrt{1 - \mathrm{Tr}(\psi U \psi U^\dagger)} = 2\sqrt{r(\mathcal{U}, \psi)} \leq 2\sqrt{r_{\mathrm{M}}(\mathcal{U})}.$$

Applying corollary 20 to the term with D gives

$$\|D\rho' D^\dagger - \rho'\|_1 \leq 1 - \sigma_{\min}(K_1)^2.$$

We now obtain an upper bound on the sum on the right-hand-side of eq. (2.38). As

$K_j \rho K_j^\dagger$ is positive semi-definite for any K_j and any $\rho \in \mathbb{D}_d$, we have

$$\begin{aligned}\sum_{j>1} \|K_j \rho K_j^\dagger\|_1 &= \sum_{j>1} \mathrm{Tr}\left(K_j \rho K_j^\dagger\right) \\ &= \mathrm{Tr}\left(\left(\sum_{j>1} K_j^\dagger K_j\right)\rho\right) \\ &= 1 - \mathrm{Tr}\left(K_1^\dagger K_1 \rho\right),\end{aligned}$$

where we have used the cyclic invariance of the trace and $\sum_j K_j^\dagger K_j = I$ for any CPTP map. As $K_1^\dagger K_1$ is positive semidefinite, we have

$$\sum_{j>1} \|K_j \rho K_j^\dagger\|_1 \leq 1 - \sigma_{\min}(K_1)^2$$

where a maximizing state is $\rho = xx^\dagger$ for any singular vector x of K_1 with singular value $\sigma_{\min}(K_1)$. Therefore we have

$$\max_{\rho \in \mathbb{D}_d} \|\mathcal{E}(\rho) - \rho\|_1 \leq 2(1 - \sigma_{\min}(K_1)^2) + \max_{\psi} 2\sqrt{r(\mathcal{U}, \psi)}. \tag{2.39}$$

As the maximum infidelity of a unitary or Hermitian matrix is stable under tensor products with the identity, the bound for $k > 1$ follows. □

Applying proposition 3 to the result gives a further upper bound in terms of the process fidelity. Though the bound is not tight for decoherent channels, their contribution to the trace distance will be linear, agreeing with theorem 4. This is implied by the appeal to corollary 20, which can be seen as a statement about the equability of a channel as proposed in [9]. Informally, this captures the idea that decoherent errors will affect subspaces uniformly as opposed to having errors that, for example, have relaxation on a single subspace but nowhere else.

For stochastic and unitary channels, however, the bound is tight; this highlights that a unitary error of the same infidelity as a stochastic or decoherent error will result in a larger diamond distance. We now demonstrate that for a family of mixed unitary channels it is also nearly saturated in an appropriate limit. For distinct Hermitian n-qubit Pauli operators P, Q, define the channel

$$\begin{aligned}\mathcal{E}_{P,Q,p,\theta}(\rho) &= p \exp(-i\theta Q)\rho \exp(i\theta Q) + (1-p)P\rho P \\ &= p(\cos^2\theta\rho + \sin^2\theta Q\rho Q - i\sin\theta\cos\theta[Q,\rho]) + (1-p)P\rho P\end{aligned} \tag{2.40}$$

where $\theta \in \mathbb{R}$ and $p \in [0,1]$. A Kraus decomposition of $\mathcal{E}_{P,Q,p,\theta}$ is

$$\begin{aligned} K_1 &= \sqrt{p}\exp(-i\theta Q) \\ K_2 &= \sqrt{1-p}P, \end{aligned} \tag{2.41}$$

and we focus on the case where K_1 is the dominant term ($p > 1/2$). This channel represents a systematic rotation around a specific axis with a small probability of a Pauli error. The polar decomposition of $K_1 = DU$ is $D = \sqrt{p}I$, $U = \exp(-i\theta Q)$ which correspond to channels $\mathcal{U}, \mathcal{D}$ with maximum infidelities

$$\begin{aligned} r_{\mathrm{M}}(\mathcal{U}) &= \sin^2\theta \\ r_{\mathrm{M}}(\mathcal{D}) &= 1-p. \end{aligned} \tag{2.42}$$

If P and Q anticommute then $R = (iPQ)^\dagger = -iQP = iPQ$ is also Hermitian and $\tau = 2^{-n}(I + iPQ)$ is a density operator with

$$\begin{aligned} 2^n(\mathcal{E}_{P,Q,p,\theta}(\tau) - \tau) &= ip\cos^2\theta PQ - ip\sin^2\theta PQ - i(1-p)PQ - iPQ - 2p\sin\theta\cos\theta Q \\ &= 2i(p - 1 - \sin^2\theta)PQ - 2p\sin\theta Q. \end{aligned} \tag{2.43}$$

By the strong unitary invariance of the Schatten norm and direct calculation

$$\begin{aligned} \|\mathcal{E}_{P,Q,p,\theta}(\tau) - \tau\|_1 &= 2^{-n}\|2i(p - 1 - \sin^2\theta)PQ - 2p\sin\theta\cos\theta Q\|_1 \\ &= 2^{-n+1}\|(1 - p + \sin^2\theta)I + ip\sin\theta\cos\theta P\|_1 \\ &= 2\sqrt{(1 - p + \sin^2\theta)^2 + p^2\sin^2\theta\cos^2\theta} \end{aligned} \tag{2.44}$$

where we have used the fact that P has 2^{n-1} eigenvalues of ± 1 each, scaling the matrix scales the eigenvalues accordingly, and adding a multiple of the identity shifts the spectrum, then noted that the singular values of the operator satisfy $\sigma_i = \sqrt{\lambda_i^* \lambda_i}$. Setting $p = 1$ or $\theta = 0$ recovers a unitary or stochastic channel from $\mathcal{E}_{P,Q,p,\theta}$ and then the bound of theorem 9 is saturated for $r_{\mathrm{M}}(\mathcal{U})$ and $r_{\mathrm{M}}(\mathcal{D})$. Furthermore, for arbitrary $p \in [0,1]$ and $\theta \in \mathbb{R}$, we have

$$\begin{aligned} (1 - p + \sin^2\theta)^2 + p^2\sin^2\theta\cos^2\theta &= \sin^2\theta(\sin^2\theta + p^2\cos^2\theta) + 2(1-p)\sin^2\theta + (1-p)^2 \\ &\leq \sin^2\theta + 2(1-p)|\sin\theta| + (1-p)^2 \\ &= ((1-p) + |\sin\theta|)^2 = (r_{\mathrm{M}}(\mathcal{D}) + \sqrt{r_{\mathrm{M}}(\mathcal{U})})^2 \end{aligned} \tag{2.45}$$

where we have used $p^2\cos^2\theta \leq 1$ and $\sin^2\theta \leq |\sin\theta|$.

A more rigorous argument can be applied by appealing to the singular values of the (at most) rank-3 operator $\mathcal{E}(\psi) - \psi$ by considering the trace formula of its characteristic polynomial, resulting in a maximization over the sensitivity to the individual unitary and stochastic errors. However, if we consider any pure state ψ that is maximally sensitive to both errors ($\{\psi, P\} = \{\psi, Q\} = 0$), we have that

$$\begin{aligned} \|\mathcal{E}_{P,Q,p,\theta}(\psi) - \psi\|_1 &= 2\|(p - 1 - \sin^2\theta)\psi - ip\sin\theta\cos\theta Q\psi\|_1 \\ &\leq 2\|(p - 1 - \sin^2\theta)I - ip\sin\theta\cos\theta Q\|_1 \end{aligned} \tag{2.46}$$

from submultiplicativity and $\|\psi\|_1 = 1$ for pure states, and has the same value as 2.44 from a similar argument.

2.4 Error scaling under unitary errors

To demonstrate the different characteristic error scalings in a quantum circuit $\mathcal{C}$ that depend on the bases of state preparation and measurement in the presence of a unitary error $\mathcal{U}$, we evaluate the total variation distance of a single qubit system with state preparation and measurement in the same basis, state preparation and measurement in different bases, and state preparation and measurement in the same basis but with a slight basis mismatch. We estimate the total variation distance τ of the probability distribution of the quantum circuit $\mathcal{C}$ to its noisy implementation $\hat{\mathcal{C}}$ under unitary error $\mathcal{U}$ by sampling

$$\begin{aligned} \tau(\mathcal{C}, \hat{\mathcal{C}}) &= \frac{1}{2}\sum_{j=0}^{1} |\langle \Pi_j, \mathcal{U}\psi\rangle - \langle \Pi_j, \psi\rangle| = |\langle \Pi_0, \mathcal{U}\psi\rangle - \langle \Pi_0, \psi\rangle| \\ &= |\mathrm{Tr}\left(\Pi_0 \left(\mathcal{U} - \mathcal{I}\right)(\psi)\right)| \end{aligned} \tag{2.47}$$

with states ψ and measurement $\Pi_0, \Pi_1 = (I - \Pi_0)$ from the Haar measure. As the Haar measure is unitarily invariant, without loss of generality the unitary channel $\mathcal{U}$ is the conjugation of U, a diagonal matrix with eigenvalues $e^{i\theta}$ and $e^{-i\theta}$.

Figure 2.1 demonstrates the difference of preparing and measuring in the same basis or different bases. When in the same basis, the total variation distance between the distributions is on the same order of magnitude as the process infidelity, while the total variation distance is on the same order of magnitude as the diamond distance when in different bases.

When there is a small basis mismatch between state preparation and measurement, the distribution of the total variation distance is not the same as when measuring and

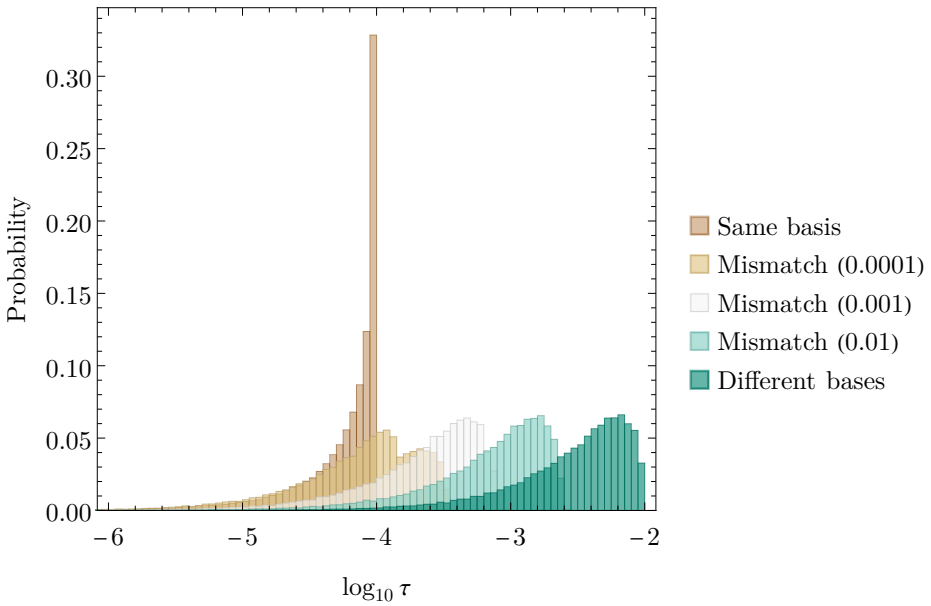

Figure 2.1: Histogram of the logarithm of the total variation distance $\tau(\mathcal{I},\mathcal{U})$ when measuring and preparing in the same basis, measuring in preparing in the same basis with a basis mismatch, and measuring and preparing in different bases. Here, the unitary U corresponding to $\mathcal{U}$ has eigenvalues $\pm e^{i\theta}$ where $\theta = 0.01$ ($r_{\mathrm{P}}(\mathcal{U}) \approx 10^{-4}$), and the basis mismatch is modelled by applying a Haar random unitary channel $\mathcal{V}$ to the PVM element while varying $r_{\mathrm{P}}(\mathcal{V})$. We have omitted varying θ as the ratio of θ to $r_{\mathrm{P}}(\mathcal{V})$ determines the behavior when both are small.

preparing in the same basis and can be magnitudes higher dependent on the severity of the mismatch. The process infidelity will underestimate the error in these cases, and dependent on the severity of said error the diamond distance could be the appropriate measure.

This can be observed more clearly as follows: for the observable $M = \vec{s}\cdot\vec{\sigma}$ and density matrix $\rho = \frac{1}{2}(I + \vec{a}\cdot\vec{\sigma})$, where $\vec{s},\vec{a} \in \mathbb{R}^3$ and $\vec{\sigma}$ is the vector of traceless Pauli matrices, the total variation distance is

$$\tau(\mathcal{I},\mathcal{U}) = |(s_x a_x + s_y a_y)\sin^2(\theta) + (s_y a_x - s_x a_y)\sin(\theta)\cos(\theta)| \tag{2.48}$$

where $\mathcal{I}$ is the identity channel. As $r_{\mathrm{P}}(\mathcal{U}) = \sin^2(\theta)$ and $\epsilon_\diamond(\mathcal{U}) = 2\sin(\theta)$, we have

$$\tau(\mathcal{I},\mathcal{U}) \leq |\vec{s}\cdot\vec{a}|\, r_{\mathrm{P}}(\mathcal{U}) + \frac{1}{2}|\vec{a}\times\vec{s}|\,\epsilon_\diamond(\mathcal{U}) \tag{2.49}$$

for small θ. While the total variation distance scales with the process infidelity when measuring and preparing in the same basis, dependent on the severity of the basis mismatch (characterized by $|\vec{a}\times\vec{s}|$) and the magnitude of the average infidelity (and thus that of the diamond distance), the diamond distance may be the dominating term for small θ. This corresponds to varying $r_{\mathrm{P}}(\mathcal{V})$ as in the example of figure 2.1, where three intermediaries between the extremes are displayed. The correct choice of measure will thus depend on both the severity of the basis mismatch and the magnitudes of the diamond distance and infidelity of $\mathcal{U}$.

2.5 Conclusion

We have obtained improved bounds on the diamond distance in terms of the infidelity and the unitarity [8], which can both be efficiently estimated. When noise is approximately stochastic, the improved bound scales as $O(r_{\mathrm{P}})$. If the unitarity indicates that the noise contains coherent errors, then the improved upper and lower bounds both scale as $O(\sqrt{r_{\mathrm{P}}})$.

Furthermore, we have shown that the non-unital contribution to the diamond distance from some generalized amplitude damping process is at most linear in the infidelity. Thus, the primary cause for the discrepancy between the infidelity and the diamond distance is due to the unital contributions of a noise process. Both of these notions are captured by considering a polar decomposition of the leading Kraus operator for the diamond distance, showing that the maximum decoherent contribution is linear as long as its singular values are not spread too far apart.

We have also shown that the diamond distance and infidelity are not "worst-case" and "average" error rates in the same sense by constructing worst-case and average versions of the infidelity and diamond distance respectively and showing that they are proportional to the standard versions up to dimensional factors. An important semantic consequence is that referring to the diamond distance as the worst-case error rate and the infidelity as the average error rate is misleading because they quantify error rates in fundamentally different ways.

By sampling the average total variation distance of a single qubit quantum circuit, we demonstrate that even small systematic errors result in a total variation distance that is qualitatively similar to state preparation and measurement in arbitrary different bases when preparing and measuring in the same basis. For deterministic computations on a noisy quantum processor, the average infidelity does not always give an average description of the error, and instead practitioners should either appeal to bounds on the diamond distance based on the unitarity, or reduce the unitary effects using randomized compiling or other error mitigation procedures.

Chapter 3

An analysis of cycle benchmarking with gate-dependent noise

Cycle benchmarking (CB) [11] is a protocol used to characterize the performance of cycles of a quantum computer. A cycle could be an individual gate or a sequence of gates run in parallel that form of a common subroutine of an algorithm. CB is a successor to randomized benchmarking in that it also samples random gates from a group and applies them between rounds of a noisy cycle to decouple cycle errors from state preparation and measurement errors. Furthermore, CB is efficient in the system size and can thus be used to characterize and probe for crosstalk and correlated errors.

While there are similarities between CB and interleaved RB [45], CB does not require the benchmarking group to contain the gate or cycle of interest. As typical interleaved RB uses Clifford gates that can each require many native gates on a platform, the reported process fidelity may not reflect actual performance of a quantum circuit or not correspond to common sequences of gates used in algorithms. By loosening this restriction, the randomizing group can be chosen with just enough structure to have a meaningful and simple fit model, while also being simple to implement and sample with a meaningful interpretation. This was done simultaneously in the context of direct RB [24] and character RB [15], where the former uses native Clifford gates to asymptotically converge to the standard RB fit model and the latter suggests an example with the use of local Clifford gates. Like randomized compiling, CB uses a group of local twirling gates that are normalized by the cycle of interest, ensuring that the noise remains approximately local — this is not true of character RB in general. While the stochastic channel learned has contributions from the implementation errors of the cycle and the randomizing gate errors, this is similar to the stochastic channels associated with the hard gates in randomized compiling [12].

In this chapter, we present a new analysis of the original Clifford CB protocol that we call simplified CB. Both protocols extract the same information and the difference is mainly aesthetic. Where the original CB protocol selected specific Pauli observables to be measured to isolate eigenvalues of the stochastic channel, simplified CB uses a character projection formula on the randomizing Pauli gates similar to the approach of character RB [15]. As the observable in standard CB is chosen to be a Pauli with the expected state after applying the circuit as a $+1$ eigenvector, this projects the output onto the specified Pauli. This is precisely what doing a character transformation to one of the randomized Pauli gates does in expectation — in this sense, they fulfill the same role. When generalizing Clifford CB to qudits using the Weyl-Heisenberg group in place of the Pauli group, the observable approach will become more complicated. Instead, using the character transform approach requires specifying an observable with non-trivial overlap with the Weyl-Heisenberg operator of interest.

Unlike character RB, simplified CB does not require the compilation and sampling of an additional gate into the final gate, which would result in requiring more circuit runs. It still however requires an additional post-processing phase to calculate the characters, but as we are dealing with the Pauli group the characters are phases and could be stored and retrieved from a lookup table with ease. The measurement of a specific Pauli observable can not generally be done natively and so standard CB needs a similar post-processing step anyway.

The main impetus for this new approach to CB is to facilitate the application of techniques developed for RB to CB, particularly the problem of robustness to gate-dependence. As there are still variations in types and severity of errors from Pauli to Pauli, the randomization may not project onto the error subspaces as desired. We show that these perturbations accumulate orthogonally to the gate-independent fit model, and that under the assumption that the noise induced by the joint implementation of the Clifford and a Pauli is close to a stochastic Pauli channel, these perturbations are small and decay more rapidly. Choosing long enough sequence lengths would then allow one to neglect the perturbations and fit to the gate independent model. We thus show that the simplified CB protocol is robust to gate-dependent noise in the same way as RB [12].

Finally, we comment on how our analysis fits in with the Fourier analysis of RB presented in [14, 46]. To this end, we show that CB can be viewed as a twisted convolution by appealing to the Cliffords as the normalizer of the Paulis, which readily fits into the perturbation theory of invariant subspaces used therein.

3.1 Preliminaries

In this section, we discuss representation theory associated with the Pauli group in terms of the PTM representation, as well as comment on the Clifford group in the same setting. While we work with qubits, we note that all of this analysis carries over to qudits by using the generalization of the Paulis to the Weyl-Heisenberg operators and using an appropriate compatible generalized Clifford group.

For non-commuting products we will use the notation $x_{a:b} = x_a x_{a-1} \dots x_{b+1} x_b$ (as well as for products $(x_i y_i)_{a:b} = x_a y_a \dots x_b y_b$) when $a \geq b$ or equal to the appropriate identity when $a < b$. Throughout this chapter we will use calligraphic font for quantum channels in the standard and Pauli-Liouville representation ($\mathcal{C}(\rho)$) and the Pauli-Liouville representation notation for unitary gates ($\phi(C)$). For general linear maps, we will also use calligraphic font (for example $\mathcal{L}$ and $\mathcal{R}$ in theorem 16).

3.1.1 Representations of the Pauli group

We begin by showing how the Pauli group decomposes in the PTM representation. For $P, Q \in \mathsf{P}_q$, we have that $PQ = \omega(P,Q)QP$, where $\omega(P,Q)$ is the phase induced by commuting the elements. We now state and prove a result about the PTM representation of the Pauli group as presented in [15].

Lemma 10. *Let P_q be the Pauli group on q qubits. The PTM decomposition for $P \in \mathsf{P}_q$ is*

$$\phi(P) = \oplus_\lambda \phi_\lambda(P) \tag{3.1}$$

where the subrepresentation ϕ_λ is 1-dimensional and has associated projector given by $P_\lambda = |\lambda\rangle\!\rangle\langle\!\langle\lambda|$, where λ are the elements of the normalized projective Pauli group.

Proof. For some $P \in \mathsf{P}_q$ and λ, we have that

$$\phi(P)|\lambda\rangle\!\rangle = |P\lambda P^\dagger\rangle\!\rangle = \omega(P,\lambda)|\lambda PP^\dagger\rangle\!\rangle = \omega(P,\lambda)|\lambda\rangle\!\rangle.$$

As P was arbitrary and $|\lambda\rangle\!\rangle$ spans a 1-dimensional subspace, $|\lambda\rangle\!\rangle$ spans an irreducible representation of ϕ which we will denote by ϕ_λ. By construction, the projector associated with ϕ_λ is $P_\lambda = |\lambda\rangle\!\rangle\langle\!\langle\lambda|$. The character associated with ϕ_λ is then

$$\chi_\lambda(P) = \mathrm{Tr}_\lambda(\phi(P)) = \mathrm{Tr}(P_\lambda \phi(P)) = \mathrm{Tr}(|\lambda\rangle\!\rangle\langle\!\langle\lambda|\phi(P)) = \langle\!\langle\lambda|P\lambda P^\dagger\rangle\!\rangle = \omega(P,\lambda),$$

which is the phase induced by commutation relations.

We now show that for $\lambda \neq \sigma$, the irreducible representations ϕ_λ and ϕ_σ are inequivalent by showing the inner product between their characters

$$\langle \chi_\lambda, \chi_\sigma \rangle = \mathbb{E}_{P \in \mathsf{P}_q} \chi_\lambda(P) \overline{\chi_\sigma(P)} = \mathbb{E}_{P \in \mathsf{P}_q} \omega(P, \lambda)\omega(P, \sigma) \tag{3.2}$$

is 0. As $\lambda \neq \sigma$, we set $\mu = \lambda\sigma \neq I$, then

$$P\mu = P\lambda\sigma = \omega(P, \lambda)\lambda P\sigma = \omega(P, \lambda)\omega(P, \sigma)\lambda\sigma P = \omega(P, \lambda)\omega(P, \sigma)\mu P,$$

implying $\omega(P, \mu) = \omega(P, \lambda)\omega(P, \sigma)$. This reduces equation 3.2 to

$$\langle \chi_\lambda, \chi_\sigma \rangle = \mathbb{E}_{P \in \mathsf{P}_q} \omega(P, \mu) = 0$$

as a non-identity Pauli will commutes with half and anti-commutes with the other half, which completes the proof. □

3.1.2 Clifford channels

We discuss the representation of ideal Clifford gates in the PTM representation. For an arbitrary Hermitian matrix P and unitary U, $(UPU^\dagger)^\dagger = UP^\dagger U^\dagger = UPU^\dagger$, and thus conjugation is Hermiticity preserving. Now, elements of the Pauli group are either Hermitian (those with phase ± 1) or anti-Hermitian (those with phase $\pm i$). As conjugation by a Clifford C permutes elements of the Pauli group and this action is bijective, each of the Hermitian basis elements of the PTM will be mapped to one other basis element up to a sign. This implies the PTM representation $\phi(C)$ of the channel $\mathcal{C}$ corresponding to C will be a signed permutation matrix.

When defining the Pauli channel learned by CB it will be useful to appeal to the orbit $\mathbb{O}(C, P) = \{C^i P C^{\dagger i}, i \geq 0\}$ of a Pauli P under the adjoint action of a Clifford C. For a Clifford of order m we have $|\mathbb{O}(C, P)| \leq m$, $|\mathbb{O}(C, P)|$ divides m, and $\mathbb{O}(C, P) = \mathbb{O}(C, Q)$ if and only if there exists j such that $P = C^j Q C^{\dagger j}$. The set of all distinct orbits is $\mathbb{O}(C)$.

3.1.3 Simplified cycle benchmarking

We present a simplified version of the CB protocol described in [11] to benchmark an implementation of a Clifford gate C. The protocol is effectively the same except that we use the character projection formula explicitly rather than estimating expectations of specific observables to induce the desired character projection.

1. Select a normalized Pauli λ.
2. Select $n+1$ random Pauli gates $\{P_i\}_{i=0}^{n}$.
3. Prepare the state ρ.
4. Apply the sequence of gates P_0, C, P_1, ..., C, P_n.
5. Measure POVM $\{Q, I-Q\}$ and multiply by the character $\overline{\chi_\lambda}(P_n C P_{n-1} \dots C P_0 C^{\dagger n})$.
6. Repeat steps 2 through 5 sufficiently many times.

The character $\chi_\lambda(P_n C P_{n-1} \dots C P_0 C^{\dagger n})$ is efficient to compute as $P_n C P_{n-1} \dots C P_0 C^{\dagger n} \in \mathsf{P}_q$ is the product of Pauli and Clifford gates, and so is simply a phase. By proposition 11 the seemingly arbitrary character is equivalent to projecting the random Pauli onto the subspace where it should be at that point in the circuit. Overall, this has the effect of acting like the inverse in standard RB in that it tracks how close to its target ideal the noisy circuit has reached.

Proposition 11. *For (normalized, projective) Pauli λ and $\vec{P} \in \mathsf{P}_q^{n+1}$, we have that*

$$\chi_\lambda(P_n C P_{n-1} \dots C P_0 C^{\dagger n}) = \prod_{i=0}^{n} \chi_{\mathcal{C}^{\dagger i}(\lambda)}(P_{n-i}). \tag{3.3}$$

Proof. Let $T = P_n C P_{n-1} \dots C P_0 C^{\dagger n}$. Then as λ is a normalized projective Pauli and $T \in \mathsf{P}_q$, we have

$$\phi(T)|\lambda\rangle\!\rangle = |T\lambda T^\dagger\rangle\!\rangle = \chi_\lambda(T)|\lambda\rangle\!\rangle. \tag{3.4}$$

We also have that

$$\begin{aligned}
\phi(T)|\lambda\rangle\!\rangle &= \phi(P_n C P_{n-1} \dots C P_0 C^{\dagger n})|\lambda\rangle\!\rangle = \phi(P_n C P_{n-1} \dots P_1 C^{\dagger n-1})\phi(C^n P_0 C^{\dagger n})|\lambda\rangle\!\rangle \\
&= \phi(P_n C P_{n-1} \dots P_1 C^{\dagger n-1})|(C^n P_0 C^{\dagger n})\lambda(C^n P_0 C^{\dagger n})^\dagger\rangle\!\rangle \\
&= \phi(P_n C P_{n-1} \dots P_1 C^{\dagger n-1})|C^n P_0 C^{\dagger n} \lambda C^n P_0^\dagger C^{\dagger n}\rangle\!\rangle \\
&= \chi_{\mathcal{C}^{\dagger n}(\lambda)}(P_0)\phi(P_n C P_{n-1} \dots P_1 C^{\dagger n-1})|\lambda\rangle\!\rangle.
\end{aligned} \tag{3.5}$$

Using a simple inductive argument on the P_i gives

$$\phi(T)|\lambda\rangle\!\rangle = \left(\prod_{i=0}^{n} \chi_{\mathcal{C}^{\dagger i}(\lambda)}(P_{n-i})\right)|\lambda\rangle\!\rangle, \tag{3.6}$$

and comparing the above with equation 3.4 gives the desired result. □

Proposition 11 can also be observed by recognizing that χ_λ is a representation of the Pauli group and the property that $\chi_\lambda(CGC^\dagger) = \chi_{C^\dagger \lambda C}(G)$ for a Clifford C.

The choices of Q and ρ are specified so that the overlaps $\langle\langle Q|\lambda\rangle\rangle$ and $\langle\langle \mathcal{C}^{\dagger n}(\lambda)|\rho\rangle\rangle$ are large — in particular, we can measure in the eigenbasis of λ and prepare a state that is a $+1$-eigenvector of $\mathcal{C}^{\dagger n}(\lambda)$. While we analyze the protocol for arbitrary n we choose $n = km$ in practice where m is the order of C acting on λ. This ensures $\mathcal{C}^{\dagger n}(\lambda) = \lambda$. We can also choose n to be the length of the orbit of λ under the adjoint action of C, as $C^n = I$ on this subspace.

Theorem 12. *The expectation of a simplified CB experiment of length n with character λ under gate-independent noise such that $\theta(G) = \mathcal{L}\phi(G)\mathcal{R}$ for $G \in \mathsf{G} = \mathsf{P}_q$ where ϕ is the PTM representation is*

$$p(n, \lambda) = A_{n,\lambda} \left(\prod_{i=1}^{n} f_{C^{\dagger i} \lambda C^i} \right) \tag{3.7}$$

where $A_{n,\lambda}$ is a SPAM constant that depends on n and λ.

Proof. The expectation over all sequences of Paulis is

$$\begin{aligned}
p(n,\lambda) &= \mathbb{E}_{\vec{P} \in \mathsf{G}^{n+1}} \overline{\chi_\lambda}(P_n C P_{n-1} \dots C P_0 C^{\dagger n}) \langle\langle Q|\theta(P_n)\theta(C)\theta(P_{n-1})\dots\theta(C)\theta(P_0)|\rho\rangle\rangle \\
&= \mathbb{E}_{\vec{T} \in \mathsf{G}^{n+1}} \overline{\chi_\lambda}(T_n) \langle\langle Q|\theta(T_n C T_{n-1}^\dagger C^\dagger)\theta(C)\theta(T_{n-1} C T_{n-2}^\dagger C^\dagger)\dots\theta(C)\theta(T_0)|\rho\rangle\rangle \\
&= \mathbb{E}_{\vec{T} \in \mathsf{G}^{n+1}} \langle\langle Q|\mathcal{L}\overline{\chi_\lambda}(T_n)\phi(T_n)\left(\phi(C)\phi(T_i)^\dagger\phi(C)^\dagger\right)\mathcal{R}\theta(C)\mathcal{L}\phi(T_i)\right)_{n-1:0}\mathcal{R}|\rho\rangle\rangle \\
&= \langle\langle Q|\mathcal{L}P_\lambda\left(\phi(C)\mathcal{D}_{\mathsf{P}_q}\right)^n\mathcal{R}|\rho\rangle\rangle
\end{aligned}$$

where we have relabelled the Paulis as $P_i \to T_i C T_{i-1}^\dagger C^\dagger$ with $T_{-1} = I$, used the gate-independent assumption on the noise and $\phi(G^\dagger) = \phi(G)^\dagger$ for the PTM representation, then applied the character projection formula and Schur's lemma with $\mathcal{D}_{\mathsf{P}_q} = \mathbb{E}_{T \in \mathsf{G}}\phi(T)^\dagger\phi(C)^\dagger\mathcal{R}\theta(C)\mathcal{L}\phi(T)$ a Pauli channel. As $P_\lambda = |\lambda\rangle\rangle\langle\langle\lambda|$ and $\langle\langle\lambda|\phi(C) = \langle\langle C^\dagger\lambda C|$,

$$P_\lambda\left(\phi(C)\mathcal{D}_{\mathsf{P}_q}\right)^n = \left(\prod_{i=1}^{n} f_{C^{\dagger i}\lambda C^i}\right)|\lambda\rangle\rangle\langle\langle C^{\dagger n}\lambda C^n|$$

where $f_\lambda = \mathrm{Tr}\left(P_\lambda\phi(C)^\dagger\mathcal{R}\theta(C)\mathcal{L}\right) = \langle\langle\lambda|\phi(C)^\dagger\mathcal{R}\theta(C)\mathcal{L}|\lambda\rangle\rangle$. Setting $A_{n,\lambda} = \langle\langle Q|\mathcal{L}_\lambda\mathcal{R}_{C^{\dagger n}\lambda C^n}|\rho\rangle\rangle$ gives the desired result. □

Corollary 13. *If $n = km$ where m is the length of the Clifford orbit containing λ, then the expectation reduces to*

$$q(k, \lambda) = A_\lambda \mu(\lambda)^k \tag{3.8}$$

where $A_\lambda = \langle\langle Q|\mathcal{L}_\lambda\mathcal{R}_\lambda|\rho\rangle\rangle$ and $\mu(\lambda) = \prod_{i=0}^{m-1} f_{C^i\lambda C^{\dagger i}}$.

By choosing $n_1 = k_1 m$ and $n_2 = k_2 m$ appropriately where m is the order of C so as to cancel the SPAM terms, we have that

$$\mu(\lambda) = \prod_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}} = \left(\frac{q(k_2, \lambda)}{q(k_1, \lambda)} \right)^{\frac{1}{k_2 - k_1}} \tag{3.9}$$

and so can amplify $\mu(\lambda)$ to achieve a multiplicative estimate [17]. The Pauli channel learned by CB is not $\mathcal{D}_{\mathsf{P}_q}$ but rather

$$\mathcal{D}_{\mathrm{CB}} = \left(\phi(C) \mathcal{D}_{\mathsf{P}_q} \right)^m = \sum_{\lambda} P_\lambda \prod_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}}, \tag{3.10}$$

where the constant $\prod_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}}$ is the same for each λ in the same orbit. In this sense, the CB protocol ambiguates all Pauli errors in a Clifford orbit — rather than being a decay rate associated to an irreducible representation as in RB, CB calculates the decay rate along an orbit. As we are interested in using this data to assess the quality of an implementation in terms of the process fidelity, we use the inequality of arithmetic and geometric means for non-negative numbers $\{x_i\}_{i=0}^{n-1}$:

$$\sqrt[n]{\prod_{i=0}^{n-1} x_i} \leq \frac{1}{n} \sum_{i=0}^{n-1} x_i. \tag{3.11}$$

We then have that

$$\begin{aligned} f_{\mathrm{P}}(\mathcal{D}_{\mathrm{CB}}^{1/m}) &= \frac{1}{4^n} \sum_{\lambda} \left(\prod_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}} \right)^{1/m} = \frac{1}{4^n} \sum_{\mathbb{O}(C)} m \left(\prod_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}} \right)^{1/m} \\ &\leq \frac{1}{4^n} \sum_{\mathbb{O}(C)} \sum_{i=0}^{m-1} f_{C^i \lambda C^{\dagger i}} = \frac{1}{4^n} \sum_{\lambda} f_\lambda = f_{\mathrm{P}}(\mathcal{D}_{\mathsf{P}_q}), \end{aligned} \tag{3.12}$$

and when $f_{\mathrm{RC}}(\mathcal{E})$ is close to 1, this bound is tight in the sense that

$$f_{\mathrm{CB}}(\mathcal{E}) - f_{\mathrm{RC}}(\mathcal{E}) = \mathcal{O}([1 - f_{\mathrm{RC}}(\mathcal{E})]^2). \tag{3.13}$$

The precision of the estimate of equation 3.9 needs to take the m^{th} root into account, however, as if we learn $\mu(\lambda)$ with an error of ε then

$$\sqrt[m]{\mu(\lambda) + \varepsilon} \approx \sqrt[m]{\mu(\lambda)} + \frac{\varepsilon (\mu(\lambda))^{\frac{1}{m} - 1}}{m}. \tag{3.14}$$

As the function $f(x) = x^{\frac{1}{m} - 1}$ goes to infinity as $x \to 0$ and $\mu(\lambda)$ is a product of m fidelities (numbers less than 1), Cliffords with large m will need a very high precision.

3.2 Cycle benchmarking is robust to gate-dependent noise

We now relax the assumption that the noise on the randomizing gates is the same across all elements of the randomizing group. While for CB the randomizing gates are Paulis and are generally well implemented, they are implemented using different physical processes. This will result in different errors from Pauli to Pauli. Rather than considering implementations of the form $\theta(G) = \mathcal{L}\phi(G)\mathcal{R}$ for linear maps $\mathcal{L}, \mathcal{R}$ and a representation $\phi(G)$, we instead construct average error maps of $\theta(G)$ in relation to a representation. We then perturb the implementation around this average implementation to show that the fit model decouples into a sum of two terms that decay exponentially, the first being the gate-independent model and the second a perturbation term. If the implementation of both the randomizing gates and the gate of interest are good[1], then the perturbation will decay rapidly.

The analysis in this section is similar to previous gate-dependent distortion theorems found in [13, 15], and relies on the fact that Cliffords normalize the Pauli group to apply a virtual inverse. Note that we do not need to further vectorize the superoperator as the irreducible representations of the Pauli group are 1-dimensional. This approach could also work when a unitary U mixes the irreps of the twirling group G in a structured way — for example, we could consider using the Clifford group as a randomizing group to benchmark a gate in the second Clifford hierarchy, which is similar to [47].

Theorem 14. *Let $\mathsf{G} = \mathsf{P}_q$ be the q qubit Pauli group with PTM representation $\phi(G) = \bigoplus_\lambda \phi_\lambda(G)$, $\theta(G)$ its corresponding (CPTP) implementation, and $\theta(C)$ a (CPTP) implementation of a Clifford C of order m. There exist linear superoperators $\mathcal{L}, \mathcal{R}$ such that*

$$\mathbb{E}_{G\in\mathsf{G}}(\phi(CG^\dagger C^\dagger)\mathcal{R}\theta(C)\theta(G)) = \phi(C)\mathcal{D}_{\mathsf{P}_q}\mathcal{R} \tag{3.15}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\theta(G)\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)) = \mathcal{L}\phi(C)\mathcal{D}_{\mathsf{P}_q} \tag{3.16}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\phi(G^\dagger C^\dagger)\mathcal{R}\theta(C)\mathcal{L}\phi(G)) = \mathcal{D}_{\mathsf{P}_q} \tag{3.17}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\phi(G)\mathcal{R}\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger)) = \phi(C)\mathcal{D}_{\mathsf{P}_q}\phi(C)^\dagger \tag{3.18}$$

where $\mathcal{D}_{\mathsf{P}_q} = \sum_\lambda f_\lambda P_\lambda$, P_λ is the projector onto the subrepresentation ϕ_λ, and f_λ chosen such that for all λ in a Clifford orbit the largest eigenvalue of the operator

$$\prod_{i=1}^{m} \left(\theta(C)\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^{\dagger i}(\lambda)}(G)^\dagger\theta(G))\right) \tag{3.19}$$

is $\prod_{i=1}^m f_{\mathcal{C}^i(\lambda)}$.

[1] As described in theorem 17.

Proof. Multiplying equations 3.15 and 3.16 from the left and right respectively by P_λ and noting that $\phi(G^\dagger) = \oplus_\lambda \phi_\lambda(G)^\dagger$ gives

$$\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^\dagger(\lambda)}(G)^\dagger|\lambda\rangle\!\rangle\langle\!\langle\lambda|\mathcal{R}\theta(C)\theta(G)) = f_{\mathcal{C}^\dagger(\lambda)}|\lambda\rangle\!\rangle\langle\!\langle\mathcal{C}^\dagger(\lambda)|\mathcal{R} \tag{3.20}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\theta(G)\theta(C)\mathcal{L}|\lambda\rangle\!\rangle\langle\!\langle\lambda|\phi_{\mathcal{C}(\lambda)}(G)^\dagger) = f_\lambda\mathcal{L}|\mathcal{C}(\lambda)\rangle\!\rangle\langle\!\langle\lambda| \tag{3.21}$$

where these equations result in systems of 4^n equations for $\mathcal{R}$ and $\mathcal{L}$. Setting $\mathcal{R} = \sum_\lambda P_\lambda\mathcal{R} = \sum_\lambda |\lambda\rangle\!\rangle\mathcal{R}_\lambda, \mathcal{L} = \sum_\lambda \mathcal{L}P_\lambda = \sum_\lambda \mathcal{L}_\lambda\langle\!\langle\lambda|$ gives

$$\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^\dagger(\lambda)}(G)^\dagger\mathcal{R}_\lambda\theta(C)\theta(G)) = f_{\mathcal{C}^\dagger(\lambda)}\mathcal{R}_{\mathcal{C}^\dagger(\lambda)} \tag{3.22}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\theta(G)\theta(C)\mathcal{L}_\lambda\phi_{\mathcal{C}(\lambda)}(G)^\dagger) = f_\lambda\mathcal{L}_{\mathcal{C}(\lambda)}. \tag{3.23}$$

As the irreducible representations ϕ_λ of P_q are 1-dimensional

$$\mathcal{R}_\lambda\theta(C)\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^\dagger(\lambda)}(G)^\dagger\theta(G)) = f_{\mathcal{C}^\dagger(\lambda)}\mathcal{R}_{\mathcal{C}^\dagger(\lambda)} \tag{3.24}$$

$$\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}(\lambda)}(G)^\dagger\theta(G))\theta(C)\mathcal{L}_\lambda = f_\lambda\mathcal{L}_{\mathcal{C}(\lambda)}. \tag{3.25}$$

We can turn the above equations into eigenvalue equations by repeatedly applying the operator of the next (previous) Pauli in the Clifford orbit of λ on the right (left) $m-1$ more times

$$\mathcal{R}_\lambda\prod_{i=1}^{m}\left(\theta(C)\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^{\dagger i}(\lambda)}(G)^\dagger\theta(G))\right) = \prod_{i=1}^{m}\left(f_{\mathcal{C}^{\dagger i}(\lambda)}\right)\mathcal{R}_\lambda \tag{3.26}$$

$$\prod_{i=1}^{m}\left(\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^{i}(\lambda)}(G)^\dagger\theta(G))\theta(C)\right)\mathcal{L}_\lambda = \prod_{i=1}^{m}\left(f_{\mathcal{C}^{i-1}(\lambda)}\right)\mathcal{L}_\lambda. \tag{3.27}$$

As $\mathcal{C}^i(\lambda) = \mathcal{C}^{\dagger m-i}(\lambda)$ for all i, the value of the scalar in the two above equations is the same. Furthermore, the above statement holds for all λ in the same Clifford orbit with the matrices shifted appropriately along the orbit, and as products of compatible matrices that are cyclically permuted have the same eigenvalues, we choose the f_λ such that $\prod_{i=1}^m f_{\mathcal{C}^i(\lambda)}$ is the largest eigenvalue of $\prod_{i=1}^m\left(\theta(C)\mathbb{E}_{G\in\mathsf{G}}(\phi_{\mathcal{C}^{\dagger i}(\lambda)}(G)^\dagger\theta(G))\right)$.

Equation 3.17 implies 3.18 by the transformation $\phi(G) \to \phi(C^\dagger G^\dagger C)$, so we focus our attention on equation 3.17. We can scale the equation by selecting Pauli channels $\mathcal{D}_R$ and $\mathcal{D}_L$, then multiplying from the left by $\phi(C)^\dagger\mathcal{D}_R\phi(C)$ and to the right by $\mathcal{D}_L$ — this is the same as setting $\mathcal{R} \to \mathcal{D}_R\mathcal{R}$ and $\mathcal{L} \to \mathcal{L}\mathcal{D}_L$. We then apply P_λ and take the trace to find

$$f_\lambda = \mathbb{E}_{G\in\mathsf{G}}(\phi_\lambda(G)^\dagger\langle\!\langle\mathcal{C}(\lambda)|\mathcal{R}\theta(C)\mathcal{L}|\lambda\rangle\!\rangle\phi_\lambda(G)) = \mathcal{R}_{\mathcal{C}(\lambda)}\theta(C)\mathcal{L}_\lambda. \tag{3.28}$$

Choosing the entires of $\mathcal{D}_R$ and $\mathcal{D}_L$ to satisfy the above equation completes the proof. □

A notion of gate-dependence is codified in the superoperators $\mathcal{L}$ and $\mathcal{R}$. All $G \in \mathsf{P}_q$ can be mapped to $\theta(G) = \mathcal{L}\phi(G)\mathcal{R} + \Delta_G$, and Δ_G can be viewed as the gate-dependence associated with G for the implementation θ. Furthermore, by the following lemma, under a convolution-like operation, $\mathcal{L}\phi(G)\mathcal{R}$ and Δ_G are orthogonal. The gate sequence that corresponds to this operation is precisely the one engineered by CB.

Lemma 15. *Let $\mathcal{L}$ and $\mathcal{R}$ be chosen as in theorem 14 for an implementation of a Clifford $\theta(C)$ and $\mathsf{G} = \mathsf{P}_q$. Then for $G_2 = GCG_1^\dagger C^\dagger$ for fixed G*

$$\mathbb{E}_{G_1 \in \mathsf{G}}(\phi(G_2)\mathcal{R}\theta(C)\Delta_{G_1}) = 0 \tag{3.29}$$
$$\mathbb{E}_{G_1 \in \mathsf{G}}(\Delta_{G_2}\theta(C)\mathcal{L}\phi(G_1)) = 0. \tag{3.30}$$

Proof. We set $\phi(G_2) \to \phi(GCG_1^\dagger C^\dagger)$ then expand the perturbation term so that the first equation is

$$\begin{aligned}
\mathbb{E}_{G_1 \in \mathsf{G}}(\phi(G_2)\mathcal{R}\theta(C)\Delta_{G_1}) &= \mathbb{E}_{G_1 \in \mathsf{G}}(\phi(GCG_1^\dagger C^\dagger)\mathcal{R}\theta(C)(\theta(G_1) - \mathcal{L}\phi(G_1)\mathcal{R})) \\
&= \phi(G)\mathbb{E}_{G_1 \in \mathsf{G}}(\phi(CG_1^\dagger C^\dagger)\mathcal{R}\theta(C)(\theta(G_1) - \mathcal{L}\phi(G_1)\mathcal{R})) \\
&= \phi(G)(\phi(C)\mathcal{D}_{\mathsf{P}_q}\mathcal{R} - \phi(C)\mathcal{D}_{\mathsf{P}_q}\mathcal{R}) \\
&= 0
\end{aligned} \tag{3.31}$$

where we have used equations 3.15 and 3.17. The second equation follows from the same argument but by applying equations 3.16 and 3.18. □

Using the above lemma, we now prove the following gate-dependent robustness result for cycle bencharmking.

Theorem 16. *The expectation value over CB experiments with $\theta(C)$ a (CPTP) implementation of a Clifford C, $\{\theta(G) : G \in \mathsf{P}_q\}$ a (CPTP) gate-dependent implementation of the Pauli group, λ a Pauli to isolate, and n the sequence length is*

$$p(n, \lambda) = A_{n,\lambda} f(n, \lambda) + \epsilon(n, \lambda) \tag{3.32}$$

where $A_{n,\lambda} = \langle\langle Q|\mathcal{L}_\lambda \mathcal{R}_{\mathcal{C}^{\dagger n}(\lambda)}|\rho\rangle\rangle$, $f(n, \lambda) = \prod_{i=1}^n f_{\mathcal{C}^{\dagger i}(\lambda)}$, and the perturbation term satisfies

$$|\epsilon(n, \lambda)| \leq \delta_1 \delta_2^n \tag{3.33}$$

where $\delta_1 = \|Q\|_1 \|\rho\|_1 \|\max_{P \in \mathsf{G}} \|\Delta_P\|_\diamond$ and $\delta_2 = \mathbb{E}_{P \in \mathsf{G}} \|\theta(C)\Delta_P\|_\diamond$.

Proof. We show that the expectation over sequences of gates $\vec{P} \in \mathsf{G}^{n+1}$ is the sum of two exponentials. Rewriting the sum over the sequence of gates with the character associated to Pauli λ we have

$$\begin{aligned} p(n,\lambda) &= \mathbb{E}_{\vec{P}\in\mathsf{G}^{n+1}} \overline{\chi_\lambda}\left(P_n (C P_i)_{n-1:0} C^{\dagger n}\right) \langle\langle Q|\theta(P_n)\left(\theta(C)\theta(P_i)\right)_{n-1:0}|\rho\rangle\rangle \\ &= \mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}} \overline{\chi_\lambda}(T_n) \langle\langle Q|\left(\theta(T_i C T_{i-1}^\dagger C^\dagger)\theta(C)\right)_{n:1}\theta(T_0)|\rho\rangle\rangle \end{aligned} \tag{3.34}$$

where $P_i = T_i C T_{i-1}^\dagger C^\dagger$ and identify $T_{-1} = I$. Expanding $\theta(T_0) = \mathcal{L}\phi(T_0)\mathcal{R} + \Delta_{T_0}$ where $\mathcal{L}$ and $\mathcal{R}$ are chosen as in theorem 14, we break the sum into two parts with

$$p(n,\lambda) = A_{n,\lambda} f(n,\lambda) + \epsilon(n,\lambda) \tag{3.35}$$

$$A_{n,\lambda} f(n,\lambda) = \mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}} \overline{\chi_\lambda}(T_n) \langle\langle Q|(\theta(T_i C T_{i-1}^\dagger C^\dagger)\theta(C))_{n:1}\mathcal{L}\phi(T_0)\mathcal{R}|\rho\rangle\rangle \tag{3.36}$$

$$\epsilon(n,\lambda) = \mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}} \overline{\chi_\lambda}(T_n) \langle\langle Q|(\theta(T_i C T_{i-1}^\dagger C^\dagger)\theta(C))_{n:1}\Delta_{T_0}|\rho\rangle\rangle. \tag{3.37}$$

From here, the idea is to expand each subsequent $\theta(P_i)$ in the expectation as $\theta(P_i) = \mathcal{L}\phi(P_i)\mathcal{R} + \Delta_{P_i}$ and use the orthogonality lemma to show that the exponentials containing ϕ and Δ are also orthogonal to one another. We first evaluate 3.36 where for $2 \le j \le n$ we have that

$$\mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}} \chi_\lambda(T_n)\left(\theta(P_i)\theta(C)\right)_{n:j+1} \Delta_{T_j C T_{j-1}^\dagger C^\dagger}\theta(C)\mathcal{L}\phi(T_{j-1}CT_{j-2}^\dagger C^\dagger)\mathcal{R}\left(\theta(C)\mathcal{L}\phi(P_i)\mathcal{R}\right)_{j-2:0} = 0 \tag{3.38}$$

by applying lemma 15 to the central term containing T_{j-1}. Using this result on equation 3.36 starting from $j = 2$ to $j = n$ we have

$$A_{n,\lambda} f(n,\lambda) = \langle\langle Q|\mathcal{L}\left(\mathbb{E}_{T_n\in\mathsf{G}}\chi_\lambda(T_n)\phi(T_n)\right)\left(\mathbb{E}_{T\in\mathsf{G}}\phi(CT^\dagger C^\dagger)\mathcal{R}\theta(C)\mathcal{L}\phi(T)\right)^n \mathcal{R}|\rho\rangle\rangle \tag{3.39}$$

$$= \langle\langle Q|\mathcal{L}P_\lambda\left(\phi(C)\mathcal{D}_{\mathsf{P}_q}\right)^n\mathcal{R}|\rho\rangle\rangle = \langle\langle Q|\mathcal{L}_\lambda\mathcal{R}_{C^{\dagger n}\lambda C^n}|\rho\rangle\rangle\left(\prod_{i=1}^n f_{C^{\dagger i}\lambda C^i}\right) \tag{3.40}$$

which is the gate-independent model obtained in theorem 12. The perturbation term 3.37 follows from the same logic — for $2 \le j \le n$ we have that

$$\mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}} \overline{\chi_\lambda}(T_n)\left(\theta(P_i)\theta(C)\right)_{n:j+1}\mathcal{L}\phi(T_j C T_{j-1}^\dagger C^\dagger)\mathcal{R}\theta(C)\Delta_{T_{j-1}CT_{j-2}^\dagger C^\dagger}\left(\theta(C)\Delta_{P_i}\right)_{j-2:0} = 0 \tag{3.41}$$

where again we apply lemma 15 to the sum over the T_{j-1} terms. This results in

$$\begin{aligned} \epsilon(n,\lambda) &= \langle\langle Q|\mathbb{E}_{\vec{T}\in\mathsf{G}^{n+1}}\overline{\chi_\lambda}(T_n)\Delta_{T_n C T_{n-1}^\dagger C^\dagger}\left(\theta(C)\Delta_{T_i C T_{i-1}^\dagger C^\dagger}\right)_{n-1:0}|\rho\rangle\rangle \\ &= \langle\langle Q|\mathbb{E}_{\vec{P}\in\mathsf{G}^{n+1}}\overline{\chi_\lambda}(P_n C P_{n-1}\ldots C P_0 C^{\dagger n})\Delta_{P_n}\left(\theta(C)\Delta_{P_i}\right)_{n-1:0}|\rho\rangle\rangle. \end{aligned} \tag{3.42}$$

Now, noting that $\chi_\lambda(G) = \pm 1$, we have that

$$\begin{aligned}|\epsilon(n,\lambda)| &\leq \mathbb{E}_{\vec{P}\in\mathsf{G}^{n+1}} |\langle\!\langle Q|\Delta_{P_n}(\theta(C)\Delta_{P_i})_{n-1:0}|\rho\rangle\!\rangle| \\ &\leq \|Q\|_1\|\rho\|_1\mathbb{E}_{\vec{P}\in\mathsf{G}^{m+1}}\|\Delta_{P_n}(\theta(C)\Delta_{P_i})_{n-1:0}\|_\diamond \\ &\leq \|Q\|_1\|\rho\|_1\|\max_{P\in\mathsf{G}}\|\Delta_P\|_\diamond \left(\mathbb{E}_{P\in\mathsf{G}}\|\theta(C)\Delta_P\|_\diamond\right)^n\end{aligned}$$

where we have used submultiplicativity of the diamond norm.

Setting $\delta_1 = \|Q\|_1\|\rho\|_1\|\max_{P\in\mathsf{G}}\|\Delta_P\|_\diamond$ and $\delta_2 = \mathbb{E}_{P\in\mathsf{G}}\|\theta(C)\Delta_P\|_\diamond$ completes the proof. □

We now show that the perturbation term is small under the assumption that the implementation is close to a representation multiplied by stochastic Pauli noise on combinations of randomizing and hard gates, providing a bound on the exponentiated portion of the perturbation term, δ_2. In this sense, the following theorem is analogous to [13, theorem 3] with $\theta(C)\theta(G)$ taking the place of $\theta(G)$. We work in a gauge where $\mathcal{R} = \mathcal{I}$ to facilitate the analysis, reducing equation 3.17 to $\mathbb{E}_{G\in\mathsf{G}}\left(\phi(G^\dagger)\theta(C)\phi(G)\right) = \mathcal{D}$ and consider perturbations of $\theta(C)\theta(G)$ around $\phi(C)\phi(G)\mathcal{D}$. While we could instead consider the gauge with $\mathcal{L} = \mathcal{I}$, due to the structure of the convolution we would need to consider perturbations of $\theta(G)\theta(C)$. This arises from having $n+1$ applications of randomly chosen $\theta(G)$ and n applications of $\theta(C)$, meaning the circuit can be considered the average over convolutions of $\theta(G)$ with $\theta_C(G) = \theta(C)\theta(G)$ n times or $\theta_C(G) = \theta(G)\theta(C)$ n times with $\theta(G)$.

Theorem 17. *Let $\mathsf{G} = \mathsf{P}_q$ and C a Clifford with associated (CPTP) implementations $\theta(G)$ and $\theta(C)$, and $\phi(U)$ the PTM representation. There exists a linear superoperator $\mathcal{L}$ satisfying theorem 14 with $\mathcal{R} = \mathcal{I}$ such that*

$$\|\theta(C)(\theta(G) - \mathcal{L}\phi(G))\| \leq \|\theta(C)\theta(G) - \phi(C)\phi(G)\mathcal{D}\| + \frac{\|\mathbb{E}_{G\in\mathsf{G}}(\theta(C)\theta(G) - \phi(C)\phi(G)\mathcal{D})\|}{1 - \|\mathcal{D}^{-1}\|\mathbb{E}_{G\in\mathsf{G}}\|\theta(C)\theta(G) - \phi(C)\phi(G)\mathcal{D}\|} \tag{3.43}$$

for any unitarily invariant norm.

Proof. As $\phi(G)$ commutes with $\mathcal{D}$ for all $G \in \mathsf{G}$, for any unitarily invariant norm we have that

$$\begin{aligned}\|\theta(C)\Delta_G\| &= \|\theta(C)\theta(G) - \theta(C)\mathcal{L}\phi(G)\| \\ &\leq \|\theta(C)\theta(G) - \phi(C)\phi(G)\mathcal{D}\| + \|\phi(C)\phi(G)\mathcal{D} - \theta(C)\mathcal{L}\phi(G)\| \\ &= \|\tilde{\mathcal{G}}\| + \|\tilde{\mathcal{C}}\|\end{aligned} \tag{3.44}$$

where $\tilde{\mathcal{G}} = \theta(C)\theta(G) - \phi(C)\phi(G)\mathcal{D}$ (which is dependent on G) and $\tilde{\mathcal{C}} = \theta(C)\mathcal{L} - \phi(C)\mathcal{D}$. Selecting $\mathcal{L}$ and $\mathcal{D}$ to satisfy theorem 14 reduces equation 3.17 to

$$\mathbb{E}_{G\in\mathsf{G}}(\phi(G^\dagger C^\dagger)\theta(C)\mathcal{L}\phi(G)) = \mathcal{D}. \tag{3.45}$$

We now consider perturbations of the equation

$$\mathbb{E}_{G\in\mathsf{G}}\left(\theta(C)\theta(G)\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right) = \theta(C)\mathcal{L}\phi(C)\mathcal{D}, \tag{3.46}$$

which is equation 3.16 left multiplied by $\theta(C)$. Expanding the right hand side gives

$$\theta(C)\mathcal{L}\phi(C)\mathcal{D} = \left(\phi(C)\mathcal{D} + \tilde{\mathcal{C}}\right)\phi(C)\mathcal{D} = (\phi(C)\mathcal{D})^2 + \tilde{\mathcal{C}}\phi(C)\mathcal{D}, \tag{3.47}$$

while expanding the left gives

$$\begin{aligned}\mathbb{E}_{G\in\mathsf{G}}\left(\theta(C)\theta(G)\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right) &= \mathbb{E}_{G\in\mathsf{G}}\left((\phi(C)\phi(G)\mathcal{D} + \tilde{\mathcal{G}})\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right)\\ &= \mathbb{E}_{G\in\mathsf{G}}\left(\phi(C)\mathcal{D}\phi(G)\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right) + \mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right)\\ &= (\phi(C)\mathcal{D})^2 + \mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right).\end{aligned} \tag{3.48}$$

Comparing equations 3.47 and 3.48, cancelling like terms, and then expanding the final $\theta(C)\mathcal{L}$ term gives

$$\begin{aligned}\tilde{\mathcal{C}}\phi(C)\mathcal{D} &= \mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\theta(C)\mathcal{L}\phi(C^\dagger G^\dagger C)\right)\\ &= \mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\phi(G^\dagger)\right)\phi(C)\mathcal{D} + \mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\tilde{\mathcal{C}}\phi(C^\dagger G^\dagger C)\right).\end{aligned} \tag{3.49}$$

Right multiplying by $\mathcal{D}^{-1}\phi(C)^\dagger$ and applying the norm gives

$$\|\tilde{\mathcal{C}}\| \leq \|\mathcal{D}^{-1}\|\|\tilde{\mathcal{C}}\|\mathbb{E}_{G\in\mathsf{G}}\|\tilde{\mathcal{G}}\| + \|\mathbb{E}_{G\in\mathsf{G}}\left(\tilde{\mathcal{G}}\phi(G^\dagger)\right)\| \tag{3.50}$$

and solving for $\|\tilde{\mathcal{C}}\|$ results in

$$\|\tilde{\mathcal{C}}\| \leq \frac{\|\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}}\phi(G^\dagger))\|}{1 - \|\mathcal{D}^{-1}\|\mathbb{E}_{G\in\mathsf{G}}\|\tilde{\mathcal{G}}\|}. \tag{3.51}$$

Inserting the upper bound on $\|\tilde{\mathcal{C}}\|$ into equation 3.44 and then expanding $\tilde{\mathcal{G}}$ gives equation 3.43. □

While the above analysis holds for any compatible unitarily invariant norm, the choice will be the diamond norm as it is compatible with theorem 16.

3.3 Fourier analysis of cycle benchmarking

In light of [14, 46], we discuss how to analyze CB as a twisted convolution of matrix functions. This approach is useful as it allows the use of a wide variety of techniques from the Fourier analysis of finite groups as developed in [48]. We define the Fourier transform $\mathcal{F}[\theta]$ of a function $\theta : \mathsf{G} \to \mathbb{M}_d$ with respect to an irreducible representation σ as

$$\mathcal{F}[\theta](\sigma) = \mathbb{E}_{G\in\mathsf{G}}\overline{\sigma}(G) \otimes \theta(G). \tag{3.52}$$

CB is based on a twisted convolution of an implementation map with itself

$$\theta *_C \theta(G) = \mathbb{E}_{H\in\mathsf{G}}\theta(GCH^\dagger C^\dagger)\theta(H) \tag{3.53}$$

where C is in the normalizer of G. We have the twisted convolution identity

$$\begin{aligned}
\mathcal{F}[\theta *_C \theta](\sigma) &= \mathbb{E}_{G\in\mathsf{G}}\overline{\sigma}(G) \otimes \theta *_C \theta(G) \\
&= \mathbb{E}_{G,H\in\mathsf{G}}\overline{\sigma}(G) \otimes \theta(GCH^\dagger C^\dagger)\theta(H) \\
&= \mathbb{E}_{G,H\in\mathsf{G}}\overline{\sigma}(GCHC^\dagger) \otimes \theta(G)\theta(H) \\
&= \mathbb{E}_{G,H\in\mathsf{G}}\overline{\sigma}(G)\overline{\sigma_{C^\dagger}}(H) \otimes \theta(G)\theta(G) \\
&= \mathbb{E}_{G\in\mathsf{G}}\overline{\sigma}(G) \otimes \theta(G)\mathbb{E}_{H\in\mathsf{G}}\overline{\sigma_{C^\dagger}} \otimes \theta(H) \\
&= \mathcal{F}[\theta](\sigma)\mathcal{F}[\theta](\sigma_{C^\dagger})
\end{aligned} \tag{3.54}$$

by using the normalized relabelling trick and the fact that σ is a representation, and then introduced the notation σ_C, the C-twisted irreducible representation of σ defined by its action on elements $G \in \mathsf{G}$ as $\sigma_C(G) = \sigma(CGC^\dagger)$. This can be extended to n-fold twisted convolution by a simple inductive argument as

$$\mathcal{F}\left[\theta^{*_C^n}\right](\sigma) = \prod_{i=0}^{n} \mathcal{F}[\theta](\sigma_{C^{\dagger i}}). \tag{3.55}$$

While this formula holds for any irreducible group representation and element of its normalizer, we restrict to the case where $\mathsf{G} = \mathsf{P}_q$ and C a Clifford (or an Abelian group and its normalizer) — this vastly simplifies the Fourier transform as the irreducible representations of P_q are 1-dimensional, so the tensor product is scalar multiplication. As the characters are isomorphic to the group elements, we see that the Fourier transform is a character projection formula onto the irreducible representations, that is

$$\mathcal{F}[\theta](\sigma) = \mathbb{E}_{G\in\mathsf{G}}\overline{\sigma}(G) \otimes \theta(G) = \mathbb{E}_{G\in\mathsf{G}}\overline{\chi_\sigma}(G)\theta(G). \tag{3.56}$$

Set $\theta(G) = \phi(CG)$ where ϕ is the PTM representation. Then the average over CB gate sequences is $\phi *_C \theta^{*_C^n}(G)$ with Fourier transform

$$\mathcal{F}\left[\phi *_C \theta^{*_C^{n-1}}\right](\sigma) = \mathcal{F}[\phi](\sigma) \prod_{i=1}^{n} \mathcal{F}[\theta](\sigma_{C^{\dagger i}}). \tag{3.57}$$

We could then apply the perturbation theory of invariant subspaces used in [46] to show that the ideal outcome of the circuit is orthogonal to perturbations in Fourier space (which is equivalent to theorem 16), and that said perturbations are small in diamond norm (equivalent to theorem 17).

3.4 Conclusion

In this chapter, we have adapted CB to the representation theory framework used in [13, 15] to prove its robustness to gate-dependent noise. We have described how the CB "decay rate" is actually the decay rate along a Clifford orbit, and how this quantity can be efficiently learned to a given precision with typical RB estimators as long as the Clifford is not of high order. This framework also makes adapting CB to qudits simple.

By considering the expectation of the randomizing gates applied between the cycle of interest, we constructed average error maps such that the implementation of the randomizing gates is the same under a convolution-like operation as a representation of the randomizing gates with the average error. Due to the convolution-like structure, the fit model decouples into the gate-independent CB decay and a perturbation term — from [14], this is a statement about the two terms being orthogonal in a Fourier space. Furthermore, we showed that if the composite error on the hard gate and randomizing gate is close to a representation multiplied by a stochastic Pauli channel, then the perturbation term decays much more rapidly than the ideal decay. This justifies neglecting the perturbation term if long enough sequences are chosen.

An open problem for the gate-dependent analysis of CB is the interpretation of the CB decay and process fidelity as they relate to RC. We touch on these topics in the following chapter.

Chapter 4

A preliminary exploration of randomized compiling with gate-dependent noise

In the previous chapter, we showed that CB with gate-dependent noise estimates the process fidelity of a composite error on an easy and hard gate. The choice of how to consider an average error on the easy gates in RC with gate-dependent noise to relate it back to this estimate is unclear, however. Perturbing around average easy gates in RC results in a stochastic channel from the approximate twirl that is different than the one estimated by CB. If instead we consider perturbations of the easy gates by inserting the average error maps constructed from CB into the RC circuit, we show with a toy model that this also results in a different stochastic channel regardless of the choice of gauge.

Gate-dependence in RB has been discussed extensively in the literature. A notion of weakly gate-dependent noise was first analyzed in [22] that derived a first-order correction to the standard RB fitting model. This approach was shown to have issues in [49] as the average gate set infidelity of a group of gates was demonstrated to have contributions not only from the noise itself but also based on the choice of representation of those gates. While the analysis of Wallman [13] settles this issue by arguing that the output of RB experiments can always be fit to a decay that corresponds to the process fidelity[1], the question of whether this number corresponds to a meaningful quantity in arbitrary circuits made of those randomizing gates remains. The CB process fidelity corresponds closely to the process fidelity of a hard gate in RC, and so we consider this question specifically as it

[1]The analysis of CB in chapter 3 does similar.

pertains to the error rate estimated by CB for use in RC when both have gate-dependent errors on their randomizing gates.

Here, we will work with a simple toy model where the easy gates are $\mathsf{E} = \mathsf{P}_q$ and the hard gates are a subset of the Clifford group $\mathbb{H} \subset \mathsf{C}_q$. For RC with gate-independent errors on the easy gates $\theta(P) = \mathcal{L}\phi(P)\mathcal{R}$, randomly compiling a circuit gives

$$\begin{aligned}
\mathcal{C}_{\mathrm{GI}}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}\theta(P_n)\theta(H_{n-1})\theta(P_{n-1})\dots\theta(H_0)\theta(P_0) \\
&= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}\mathcal{L}\phi(P_n)\mathcal{R}\theta(H_{n-1})\mathcal{L}\phi(P_{n-1})\mathcal{R}\dots\theta(H_0)\mathcal{L}\phi(P_0)\mathcal{R} \\
&= \mathcal{L}(\phi(H_i)\mathbb{E}_{P\in\mathsf{P}_q}(\phi(P)^\dagger\phi(H_i)^\dagger\mathcal{R}\theta(H_i)\mathcal{L}\phi(P))_{n-1:0}\mathcal{R} \\
&= \mathcal{L}\phi(H_{n-1})\mathcal{D}_{\mathsf{P}_q}(H_{n-1})\dots\phi(H_0)\mathcal{D}_{\mathsf{P}_q}(H_0)\mathcal{R} \qquad (4.1)
\end{aligned}$$

by a simple application of Schur's lemma, where $\vec{H} \in \mathbb{H}^n$ is a sequence of hard gates and $\mathcal{D}_{\mathsf{P}_q}(H_i) = \mathbb{E}_{P\in\mathsf{P}_q}(\phi(P)\phi(H_i)^\dagger\mathcal{R}\theta(H_i)\mathcal{L}\phi(P)^\dagger)$. As the linear maps $\mathcal{L}, \mathcal{R}$ are independent of the randomizing gate and the hard gates, the stochastic channels $\mathcal{D}_{\mathsf{P}_q}$ align with the approximation from CB as shown by theorem 12. This implies gate-independent CB gives an accurate estimation of the process fidelity of a hard gate in gate-independent RC.

Moving from gate-independent to gate-dependent noise on the randomizing gates makes this interpretation much less clear, however, as the average error maps used to prove CB's robustness to gate-dependence in chapter 3 not only depend on the randomizing gates but also on the gate or cycle they are characterizing. At each step in CB, we are amplifying the same composite error which gives an operational understanding to the perturbation analysis: we are perturbing around the average composite error of randomized gates and a hard gate. On the other hand, with randomized compiling we consider arbitrary circuits with different hard gates at each step, and so there is no single composite easy-hard error from which to perturb.

This can be illustrated by considering this as a problem of gauge. We can represent a quantum information processor as a group of quantum gates G with corresponding implementations $\theta(\mathsf{G})$, a set of states $\{\rho_i\}$, and a POVM $\{F_i\}$ [49, 50]. In the PTM representation, given an initial state ρ_j, a sequence of gates $\vec{G} \in \mathsf{G}^n$, the Born rule gives that the probability of measuring F_i is

$$p(i|\vec{G}, \rho_j) = \langle\langle F_i|\theta(G_i)_{n:1}|\rho_j\rangle\rangle. \qquad (4.2)$$

Given an invertible linear map M, the representation

$$\langle\langle F_i| \to \langle\langle F_i|M^{-1}, \quad \theta(\mathsf{G}) \to M\theta(\mathsf{G})M^{-1}, \quad |\rho_j\rangle\rangle \to M|\rho_j\rangle\rangle \qquad (4.3)$$

preserves the probabilities of equation 4.2. This type of transformation is called a gauge transformation [49, 50]. While the probabilities will be conserved, note that the choice of M could lead to a situation where $\theta(\mathsf{G})$ are not CPTP maps, $\{\rho_i\}$ are not density matrices, or $\{F_i\}$ is not a POVM, but we will not pursue this issue further.

Given an implementation of a Clifford $\theta(H)$, under gate-dependent noise on the randomizing gates, CB learns the noise associated with the operator $\phi(H)^\dagger \mathcal{R}_{\text{CB}} \theta(H) \mathcal{L}_{\text{CB}}$ where $\mathcal{R}_{\text{CB}}, \mathcal{L}_{\text{CB}}$ are the average error maps constructed from theorem 14 with $\theta(H)$. By theorem 17, in the gauge where $\mathcal{R}_{\text{CB}} = \mathcal{I}$ the perturbation term $\Delta_G(H)$ is small for each G in the randomizing group. This gauge is obtained by setting $M = \mathcal{R}_{\text{CB}}^{-1}$[2], and we could equally consider inverting $\mathcal{L}_{\text{CB}}$ instead. In this setting, we are amplifying the average error associated with implementations

$$\theta(H)\theta(G) = \theta(H)\mathcal{L}_{\text{CB}}\phi(G) + \theta(H)\Delta_G(H) \tag{4.4}$$

by inserting a virtual inverse with the normalized relabelling trick.

For RC, the choice of average error maps in the presence of gate-dependence is less clear. In view of the gate-independent model, we would like to represent the randomizing gates as

$$\theta(G) = \mathcal{L}_{\text{RC}}\phi(G)\mathcal{R}_{\text{RC}} + \Delta_G \tag{4.5}$$

where the error maps have no dependence on hard gates, and we could again select a gauge transformation to eliminate either the left or right error map. By applying theorem 14 without a hard gate (equivalently, setting $\theta(H) = \mathcal{I}$) we can construct error maps that corresponds to the noise between the Pauli gates, $\mathcal{R}_{\text{RC}}\mathcal{L}_{\text{RC}}$, which for the toy model is the natural choice as it captures information across all the Paulis. If inserted into the RC circuit the noise between gates is

$$\phi(H_i)^\dagger \mathcal{R}_{\text{RC}} \theta(H_i) \mathcal{L}_{\text{RC}} \neq \phi(H_i)^\dagger \mathcal{R}_{\text{CB}} \theta(H_i) \mathcal{L}_{\text{CB}}. \tag{4.6}$$

This also prevents us from availing of the orthogonality lemma (lemma 15).

An alternative choice is to select a different set of average noise operators for each Pauli that correspond to the hard gate on the left, right, or the corresponding error map for the hard gate on either side. By the following theorem, however, this is not a good choice.

Theorem 18. *Given a desired unitary U that can be decomposed into Clifford gates $\vec{H} \in \mathsf{C}_q^n$ with implementations $\theta(H_i)$, the expectation when randomly compiling in a CPTP imple-*

[2] If $\mathcal{R}_{\text{CB}}$ is not invertible, it can always be made invertible with an arbitrarily small perturbation

mentation of the Pauli gates $\{\theta(P_i\})$ *is*

$$\begin{aligned}\mathcal{C}_{GD}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_n)\theta(H_{n-1})\theta(P_{n-1})\dots\theta(H_0)\theta(P_0)) \\ &= \mathcal{L}_n(\phi(H_i)\widetilde{\mathcal{D}}_{\mathsf{P_q}}(H_i))_{n-1:0}\mathcal{R}_0 + \sum_{j=0}^{n-1}\delta_j(\vec{H})\end{aligned} \tag{4.7}$$

where

$$\widetilde{\mathcal{D}}_{\mathsf{P}_q}(H_i) = \mathbb{E}_{P\in\mathsf{P}_q}\phi(P)\phi(H_i)^\dagger\mathcal{R}_{i+1}\theta(H_i)\mathcal{L}_i\phi(P)^\dagger, \tag{4.8}$$

$$\delta_j(\vec{H}) = \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+2}\Delta^j_{P_{j+1}}\theta(H_j)\Delta^j_{P_j}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j-1:0}, \tag{4.9}$$

$\mathcal{L}_i$,$\mathcal{R}_i$, *and* Δ^i_P *are the gate-dependent left and right error maps and gate-dependent perturbation for* $\theta(H_i)$ *respectively, with* $\mathcal{L}_n = \mathcal{L}_{n-1}$, $\mathcal{R}_n = \mathcal{R}_{n-1}$.

Proof. We sequentially expand the $\theta(P_i) = \mathcal{L}_i\phi(P_i)\mathcal{R}_i + \Delta^i_{P_i}$ from $i = 0$ to $n-1$ by using the error maps $\mathcal{L}_i$ and $\mathcal{R}_i$ constructed for $\theta(H_i)$ in theorem 14 respectively. In particular, this gives

$$\begin{aligned}\mathcal{C}_{\mathrm{GD}}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}\theta(P_n)\theta(H_{n-1})\theta(P_{n-1})\dots\theta(H_0)(\mathcal{L}_0\phi(P_0)\mathcal{R}_0 + \Delta^0_{P_0}) \\ &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:1}\mathcal{L}_0\phi(P_0)\mathcal{R}_0 + \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:1}\Delta^0_{P_0} \\ &= \mathcal{C}_0(\vec{H}) + \delta_0(\vec{H})\end{aligned} \tag{4.10}$$

This process proceeds on for all $j = 1\dots n-1$ by expanding $\theta(P_{j+1})$

$$\begin{aligned}\mathcal{C}_j(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+1}\theta(P_{j+1})(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j:0} \\ &= \mathcal{C}_{j+1}(\vec{H}) + \delta_{j+1}(\vec{H}) \\ \mathcal{C}_{j+1}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+2}\mathcal{L}_{j+1}\phi(P_{j+1})\mathcal{R}_{j+1}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j:0}, \\ \delta_{j+1}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+2}\Delta^{j+1}_{P_{j+1}}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j:0}.\end{aligned} \tag{4.11}$$

The randomly compiled circuit is then

$$\mathcal{C}_{\mathrm{GD}}(\vec{H}) = \mathcal{C}_{n-1}(\vec{H}) + \sum_{i=0}^{n-1}\delta_i(\vec{H}), \tag{4.12}$$

where we then expand the edge term $\theta(P_n) = \mathcal{L}_n\phi(P_n)\mathcal{R}_n + \Delta^n_{P_n}$ but choosing $\mathcal{L}_n$, $\mathcal{R}_n$ from theorem 14 with $\theta(H_{n-1})$, apply lemma 15 to remove the final perturbation term, then sum over the Paulis with the transformation $P_{i+1} = T_{i+1}H_iT_i^\dagger H_i^\dagger$

$$\begin{aligned}\mathcal{C}_{n-1}(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}\theta(P_n)\theta(H_{n-1})\mathcal{L}_{n-1}\phi(P_{n-1})\mathcal{R}_{n-1}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{n-2:0}\\ &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}\mathcal{L}_n\theta(P_n)\mathcal{R}_n(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{n-1:0}\\ &= \mathcal{L}_n(\phi(H_i)\widetilde{\mathcal{D}}_{\mathsf{P_q}}(H_i))_{n-1:0}\mathcal{R}_0. \end{aligned} \tag{4.13}$$

Finally, the perturbation terms can shown to be second order by considering a different perturbation point. For δ_j we can expand $\theta(P_{j+1})$ around the maps constructed for $\theta(H_j)$ to get

$$\begin{aligned}\delta_j(\vec{H}) &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+2}(\mathcal{L}_j\phi(P_j)\mathcal{R}_j + \Delta^j_{P_{j+1}})\theta(H_j)\Delta^j_{P_j}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j-1:0}\\ &= \mathbb{E}_{\vec{P}\in\mathsf{P}_q^{n+1}}(\theta(P_i)\theta(H_{i-1}))_{n:j+2}\Delta^j_{P_{j+1}}\theta(H_j)\Delta^j_{P_j}(\theta(H_i)\mathcal{L}_i\phi(P_i)\mathcal{R}_i)_{j-1:0}, \end{aligned} \tag{4.14}$$

completing the proof. □

The stochastic channels $\widetilde{\mathcal{D}}_{\mathsf{P}_q}$ in theorem 18 now have a dependence on error maps from distinct hard gates which is certainly not well estimated by CB, and furthermore, as the $\mathcal{L}_i$ and $\mathcal{R}_i$ for each $\theta(H_i)$ are distinct, there is no gauge transformation to remove all of their dependence simultaneously. Similarly, setting $\mathcal{L}_i, \mathcal{R}_i$ according to theorem 14 for $\theta(H_{i+1}), \theta(H_i)$ respectively ensures that the channels align with the CB channels, but at the cost of having a meaningful relation to the perturbation term and orthogonality lemma. We note that the perturbation terms are second order in the Δ^i_G, however, implying there is some error suppression.

To remedy this situation, a starting point is to consider the RC error maps $\mathcal{L}_{\text{RC}}, \mathcal{R}_{\text{RC}}$ as perturbations from the CB error maps $\mathcal{L}_{\text{CB}}, \mathcal{R}_{\text{CB}}$ at each step (or vice versa), as this would allow the use of an approximate orthogonality relation to track and suppress the evolution of these perturbations. Ideally, this would lead to a stability theorem for terms of the form $\|\mathcal{L}_{\text{CB}} - \mathcal{L}_{\text{RC}}\|_\diamond, \|\mathcal{R}_{\text{CB}} - \mathcal{R}_{\text{RC}}\|_\diamond$. While we do not expect to see a full decoupling of the perturbation term from the fit model, this will likely result in a proper interpretation for the gate-dependent process fidelity.

Moving back to a universal gate set from the toy model further complicates the problem. In addition to obfuscating easy and hard errors, the easy gates will now have non-Paulis or a fixed Pauli compiled in — while the number of different possible easy gates at each step will not change, the maps $\mathcal{L}_{\text{RC}}, \mathcal{R}_{\text{RC}}$ will also depend on this choice which could be even further away from the CB average error maps.

We posit that this is a mathematical issue to be resolved as opposed to being indicative of a wider disconnect between the two protocols. The RB process fidelity is proportional to the exponential increase in the probability of error when increasing the depth of a circuit as described by the gate set circuit fidelity [51]. We believe that the CB process fidelity captures the same information about RC circuits — this aligns with what has been observed in RC experiments [52].

Chapter 5

Conclusion and Future Work

In this thesis, we began by considering the appropriate, context-dependent use of distance measures for quantum channels. Besides the semantic consequences implied by the bounds contained in chapter 2, we also demonstrate the destructive nature of coherent errors in quantum circuits, as well as the (relative) benignity of stochastic and decoherent errors. This highlights the necessity of error mitigation techniques for coherent errors like randomized compiling. A similar approach to analyzing the total variation distance in section 2.48 with synthesized unitaries could be used to explore the effectiveness of these mitigation protocols.

Many of the bounds considered therein do not have ideal dimensional scaling. The space of all CPTP maps is not representative of small implementation error maps, and so restricting our attention based on reasonable physical assumptions can lead to tighter bounds in terms of dimension. As many quantum computers do not have arbitrary couplings between constituent systems, arbitrary bleed-through of pulses is not likely to occur. Considering channels generated by k-local pulses would facilitate this.

We then proved that CB is robust to gate-dependent noise. A natural extension of CB that we are currently developing is a protocol to learn the exact stochastic channel associated with an implementation of a cycle, which provides a intermediary between RB and gate set tomography. As CB learns an approximation to the actual channel that arises in RC that lower bounds the process fidelity, it would be beneficial for applications like error mitigation and control to have more fine-grained information about the channel. Although this has already been developed and shown to be efficient under specific sparsity conditions on the noise [53] for implementations of the Pauli group without interleaved gates, the degeneracy caused by the Clifford orbits in CB would need to be broken to learn

this information for an interleaved gate.

By extending the group of randomizing gates between application of the Clifford gate to include deterministic gates to accumulate errors outside of the subspaces determined by the Clifford orbits, we can learn a system of channels to isolate or approximate individual Pauli fidelities using only local gates. The gate-dependent analysis for CB is immediately inherited by said approach, as well as CB for non-Clifford gates with some minor adjustments.

While there are gate-dependent results presented in the original randomized compiling paper [8], we plan to apply this type of analysis coupled with the Fourier analysis developed in [46]. Finally, we have initiated a discussion about the relationship of the gate-dependent process fidelity in CB to the stochastic channels in RC. Connecting the two is a non-trivial problem, especially when considering a randomly compiled universal gate set.

References

[1] Seth Lloyd. Universal Quantum Simulators. *Science*, 273(5278):1073, aug 1996.

[2] Katherine L. Brown, William J. Munro, and Vivien M. Kendon. Using quantum computers for quantum simulation. *Entropy*, 12(11):2268–2307, Nov 2010.

[3] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Review*, 41(2):303, 1999.

[4] Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation, 2009.

[5] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Physical Review A*, 77(3), Mar 2008.

[6] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nature Communications*, 8(1), Feb 2017.

[7] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B*, 7(10):S347, 2005.

[8] Joel J. Wallman, Christopher Granade, Robin Harper, and Steven T. Flammia. Estimating the Coherence of Noise. *New Journal of Physics*, 17:113020, 2015.

[9] Arnaud Carignan-Dugas, Matthew Alexander, and Joseph Emerson. A polar decomposition for quantum channels (with applications to bounding error propagation in quantum circuits). *Quantum*, 3:173, August 2019.

[10] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, may 1999.

[11] Alexander Erhard, Joel James Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban Adrian Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *Nature Communications*, 2019.

[12] Joel J. Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A*, 94:052325, 2016.

[13] Joel J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, 2018.

[14] Seth T. Merkel, Emily J. Pritchett, and Bryan H. Fong. Randomized benchmarking as convolution: Fourier analysis of gate dependent errors, 2019.

[15] Jonas Helsen, Xiao Xue, Lieven M. K. Vandersypen, and Stephanie Wehner. A new class of efficient randomized benchmarking protocols. *npj Quantum Information*, 5(1):71, Aug 2019.

[16] Christopher Granade, Christopher Ferrie, and D G Cory. Accelerated randomized benchmarking. *New Journal of Physics*, 17(1):013042, Jan 2015.

[17] Robin Harper, Ian Hincks, Chris Ferrie, Steven T. Flammia, and Joel J. Wallman. Statistical analysis of randomized benchmarking. *Phys. Rev. A*, 99(5):052350, may 2019.

[18] Mahnaz Jafarzadeh, Ya-Dong Wu, Yuval R Sanders, and Barry C Sanders. Randomized benchmarking for qudit clifford gates. *New Journal of Physics*, 22(6):063014, Jun 2020.

[19] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.

[20] Alexei Gilchrist, Nathan Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, jun 2005.

[21] Michael A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249, oct 2002.

[22] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85:042311, 2012.

[23] W. Fulton and J. Harris. Representation theory: A first course. In *Representation Theory: A First Course*, 1991.

[24] Timothy J. Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Direct randomized benchmarking for multiqubit devices. *Physical Review Letters*, 123(3), Jul 2019.

[25] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, John A. Smolin, Jerry M. Chow, Colm A. Ryan, Chad Rigetti, S. Poletto, Thomas A. Ohki, and et al. Characterization of addressability by simultaneous randomized benchmarking. *Physical Review Letters*, 109(24), Dec 2012.

[26] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. Bounding the average gate fidelity of composite channels using the unitarity. *New Journal of Physics*, 21(5):053016, May 2019.

[27] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. Characterizing universal gate sets via dihedral benchmarking. *Phys. Rev. A*, 92:060302, Dec 2015.

[28] O. Kern, G. Alber, and D. L. Shepelyansky. Quantum error correction of coherent errors by randomization. *The European Physical Journal D*, 32(1):153–156, Jan 2005.

[29] Lea F Santos and Lorenza Viola. Advantages of randomization in coherent quantum dynamical control. *New Journal of Physics*, 10(8):083009, Aug 2008.

[30] Earl Campbell. Shorter gate sequences for quantum computing by mixing unitaries. *Physical Review A*, 95(4), Apr 2017.

[31] Earl Campbell. Random compiler for fast hamiltonian simulation. *Physical Review Letters*, 123(7), Aug 2019.

[32] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):2493, 1995.

[33] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM J. Comput.*, 38(4):1207–1282, July 2008.

[34] Emanuel Knill, Raymond Laflamme, and Wojciech Hubert Zurek. Resilient Quantum Computation. *Science*, 279(5349):342–345, jan 1998.

[35] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80:012304, 2009.

[36] Emanuel Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, jan 2008.

[37] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Scalable and Robust Randomized Benchmarking of Quantum Processes. *Physical Review Letters*, 106:180504, 2011.

[38] Joel J. Wallman and Steven T. Flammia. Randomized benchmarking with confidence. *New Journal of Physics*, 16(10):103032, oct 2014.

[39] Yuval R Sanders, Joel J. Wallman, and Barry C. Sanders. Bounding quantum gate error rate based on reported average fidelity. *New Journal of Physics*, 18(1):012002, dec 2015.

[40] Lukasz Cincio, Kenneth Rudinger, Mohan Sarovar, and Patrick J. Coles. Machine learning of noise-resilient quantum circuits. *arXiv ePrint*, 2020.

[41] Richard Kueng, David M. Long, Andrew C. Doherty, and Steven T. Flammia. Comparing Experiments to the Fault-Tolerance Threshold. *Physical Review Letters*, 117(17):170502, oct 2016.

[42] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, Jan 1998.

[43] Holger F. Hofmann. Complementary Classical Fidelities as an Efficient Criterion for the Evaluation of Experimentally Realized Quantum Operations. *Physical Review Letters*, 94, 2005.

[44] John Watrous. Simpler semidefinite programs for completely bounded norms. *Chicago Journal of Theoretical Computer Science*, 2012.

[45] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, and et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Physical Review Letters*, 109(8), Aug 2012.

[46] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H. Werner, and Jens Eisert. A general framework for randomized benchmarking, 2020.

[47] Robin Harper and Steven T Flammia. Estimating the fidelity of t gates using standard interleaved randomized benchmarking. *Quantum Science and Technology*, 2(1):015008, Mar 2017.

[48] W. T. Gowers and O. Hatami. Inverse and stability theorems for approximate representations of finite groups, 2016.

[49] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout. What randomized benchmarking actually measures. *Physical Review Letters*, 119(13), Sep 2017.

[50] Junan Lin, Brandon Buonacorsi, Raymond Laflamme, and Joel J Wallman. On the freedom in representing quantum operations. *New Journal of Physics*, 21(2):023006, Feb 2019.

[51] Arnaud Carignan-Dugas, Kristine Boone, Joel J Wallman, and Joseph Emerson. From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters. *New Journal of Physics*, 20(9):092001, Sep 2018.

[52] Akel Hashim, Ravi K. Naik, Alexis Morvan, Jean-Loup Ville, Bradley Mitchell, John Mark Kreikebaum, Marc Davis, Ethan Smith, Costin Iancu, Kevin P. O'Brien, Ian Hincks, Joel J. Wallman, Joseph Emerson, and Irfan Siddiqi. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, 2020.

[53] Steven T. Flammia and Joel J. Wallman. Efficient estimation of pauli channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, Dec 2020.

[54] Kristine Boone, Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. Randomized benchmarking under different gate sets. *Physical Review A*, 99(3), Mar 2019.

[55] Peter Selinger. Generators and relations for n-qubit clifford operators. *Logical Methods in Computer Science*, 11(2), Jun 2015.

APPENDICES

Appendix A

A bound on differences of Kraus operators

In this appendix, we prove a bound on terms that appear when taking differences between CPTP maps. We begin with a general inequality and then use it to bound the difference between a 'decoherent' Kraus operator and the identity channel.

Proposition 19. *For any operators A, B, and C,*

$$\|ACA^\dagger - BCB^\dagger\|_1 \leq \|A-B\|_\infty \|A+B\|_\infty \|C\|_1.$$

Proof. First note that

$$ACA^\dagger - BCB^\dagger = \frac{1}{2}\left((A+B)C(A-B)^\dagger + (A-B)C(A+B)^\dagger\right).$$

Therefore by the triangle inequality, we have

$$\|ACA^\dagger - BCB^\dagger\|_1 \leq \tfrac{1}{2}\|(A+B)C(A-B)^\dagger\|_1 + \tfrac{1}{2}\|(A-B)C(A+B)^\dagger\|_1.$$

As the Schatten 1-norm is Hölder-dual to the Schatten ∞-norm, we have

$$\begin{aligned}\|(A+B)C(A-B)^\dagger\|_1 &\leq \|A+B\|_\infty \|C(A-B)^\dagger\|_1 \\ &\leq \|A+B\|_\infty \|A-B\|_\infty \|C\|_1,\end{aligned}$$

where we have used $\|M^\dagger\|_\infty = \|M\|_\infty$. The same argument holds for $\|(A-B)C(A+B)^\dagger\|_1$, proving the general statement. □

Corollary 20. *Let $K \in \mathbb{C}^{d\times d}$ be a positive operator such that $\sigma_{\max}(K) \leq \min\{3\sigma_{\min}(K), 1\}$. Then*

$$\max_{\rho\in\mathbb{D}_d} \|K\rho K^\dagger - \rho\|_1 = 1 - \sigma_{\min}(K)^2.$$

Proof. By the triangle inequality,

$$\begin{aligned}\|K\rho K^\dagger - \rho\|_1 &\leq \|K\rho K^\dagger - \lambda^2\rho\|_1 + \|(\lambda^2 - 1)\rho\|_1 \\ &\leq \|K\rho K^\dagger - \lambda^2\rho\|_1 + 1 - \lambda^2.\end{aligned}$$

By theorem 19 with $A = K$, $B = \lambda I$ and $C = \rho \in \mathbb{D}_d$,

$$\|K\rho K^\dagger - \lambda^2\rho\|_1 \leq \|K - \lambda I\|_\infty \|K + \lambda I\|_\infty.$$

Choosing $\lambda \leq (\sigma_{\min}(K) + \sigma_{\max}(K))/2$ so that the eigenvalue of $K - \lambda I$ with the largest modulus is positive, we have

$$\|K\rho K^\dagger - \lambda^2\rho\|_1 \leq (\sigma_{\max}(K) - \lambda)(\sigma_{\max}(K) + \lambda) = \sigma_{\max}(K)^2 - \lambda^2.$$

and so

$$\|K\rho K^\dagger - \rho\|_1 \leq 1 + \sigma_{\max}(K)^2 - 2\lambda^2.$$

Setting

$$\begin{aligned}\lambda &= \frac{\sqrt{\sigma_{\max}(K)^2 - \sigma_{\min}(K)^2}}{\sqrt{2}} \\ &= \frac{\sigma_{\min}(K) + \sigma_{\max}(K)}{2}\sqrt{2}\sqrt{\frac{\sigma_{\max}(K) - \sigma_{\min}(K)}{\sigma_{\max}(K) + \sigma_{\min}(K)}}\end{aligned}$$

completes the proof, provided $\lambda \leq (\sigma_{\min}(K) + \sigma_{\max}(K))/2$, that is, provided

$$\sigma_{\max}(K) - 3\sigma_{\min}(K) \leq 0.$$

□